Learning Ideological Latent space in Twitter

Preethi Lahoti

School of Science

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo July 31, 2017

Thesis supervisor:

Prof. Aristides Gionis

Thesis advisor:

M.Sc. Kiran Garimella

Aalto University
School of Science

AALTO UNIVERSITY
SCHOOL OF SCIENCE

ABSTRACT OF THE
MASTER'S THESIS

Author: Preethi Lahoti

Title: Learning Ideological Latent space in Twitter

Date: July 31, 2017 Language: English Number of pages: 5+47

Department of Computer Science

Professorship:

Supervisor: Prof. Aristides Gionis

Advisor: M.Sc. Kiran Garimella

People are shifting from traditional news sources to online news at an incredibly fast rate. However, the technology behind online news consumption forces users to be confined to content that confirms with their own point of view. This has led to social phenomena like polarization of point-of-view and intolerance towards opposing views. In this thesis we study information filter bubbles from a mathematical standpoint. We use data mining techniques to learn a *liberal-conservative* ideology space in Twitter and presents a case study on how such a latent space can be used to tackle the filter bubble problem on social networks.

We model the problem of learning *liberal-conservative* ideology as a constrained optimization problem. Using matrix factorization we uncover an ideological latent space for content consumption and social interaction habits of users in Twitter. We validate our model on real world Twitter dataset on three controversial topics - "obamacare", "gun control" and "abortion". Using the proposed technique we are able to separate users by their ideology with 95% *purity*. Our analysis shows that there is a very high correlation $(0.8 - 0.9)$ between the estimated ideology using machine learning and true ideology collected from various sources.

Finally, we re-examine the learnt latent space, and present a case study showcasing how this ideological latent space can be used to develop exploratory and interactive interfaces that can help in diffusing the information filter bubble. Our matrix factorization based model for learning ideology latent space, along with the case studies provide a theoretically solid as well as a practical and interesting point-of-view to online polarization. Further, it provides a strong foundation and suggests several avenues for future work in multiple emerging interdisciplinary research areas, for instance, humanly interpretable and explanatory machine learning, transparent recommendations and a new field that we coin as *Next Generation Social Networks*.

Keywords: Information Filter Bubble, Social Media, Twitter, Polarization, Matrix Factorization, Combining Link and Content, Ideology Score, Latent Space Learning, Graph Regularization

Preface

This thesis would not have been possible without the support by the Aalto University's Computer Science department, Tekes project "Re:Know," and the EC H2020 project "SoBigData" (654024).

I would like to express my deepest gratitude to my thesis supervisor Prof. Aristides Gionis who has taught me all I know about Data Mining. I would like to thank my thesis advisor Kiran Garimella, this thesis would not have been possible without his excellent guidance throughout the work. I would like to thank all my teachers and professors. I would like to additionally thank all my colleagues for engaging discussions and insightful comments that gave inspiration to many ideas during the thesis. I would like to thank my university staff for their support throughout my thesis.

Most of all, I would like to thank my dear husband Robert for his incessant support, great ideas and exciting discussions. His presence and the means provided, made the everyday routine significantly more luxurious and enjoyable.

Espoo, July 31, 2017

Preethi Lahoti

Contents

1 Introduction

1.1 Motivation and Scope

Social media and the web was envisioned with the goal to encourage and provide diverse and unreachable information around the world to users with ease. However, over the past few years, various factors such as design of social networks, algorithmic filtering of content (e.g., news feeds, recommendations), have narrowed the content that a user sees and consumes. Recent studies on US presidential elections have shown that the content viewed by Republican and Democratic voters was very different. For instance, imagine two users of opposite ideological stances (say *liberal* and *conservative*). The two users may be looking at the content on the same topic (*presidential debate*), however what they see can be of completely different viewpoints. Consequently, each user lives in their own information bubble, oblivious to the views on the other side, creating their own world-view of truth. This has led to social phenomena like polarization of point-of-view and intolerance towards opposing views. Especially for controversial topics, this phenomenon has led to ideological segregation. Studies suggest that over the years users will increasingly live in their echo chambers [31]. In the book Sunstein [35] the author discusses the threat that these phenomenon will pose to open democratic processes. Recent events like US Elections and BREXIT are excellent examples of it.

Here after we refer to this phenomena where users get less exposure to conflicting viewpoints and are intellectually isolated in their own informational bubble as “filter bubble”, a term coined by Pariser [29]. Further, we call this tendency of a user to view content of certain viewpoint more than the other viewpoint as bias. The cause for this bias can be two folds (i) The user by himself prefers content from one viewpoint over another, i.e., “user bias” (ii) various set of algorithms on web, for instance, recommender, structure of social network, search engines together impose the filter on the user, i.e., “imposed bias”. Let us elaborate with an example. Consider a user who is interested in a certain topic exploring the web. The content that the user sees/consumes is hugely influenced by a variety of online systems and technologies such as social networks, recommendation systems and information retrieval engines. (i) Social networks - The presence of “homophily” in user’s social interactions and social links, when combined with selectively filtered social feed from ideologically similar users would largely expose the user only to the content that confirms with his ideology (ii) Recommender systems - Recommender systems optimize to suggest content that a user is likely to consume based on the history of the user, a criteria that further contributes to the filter bubble. As a consequence, recommendation systems that the user would be exposed to would further recommend content that agrees with user’s historical viewpoint. (iii) Likewise, search engines preferentially rank results based on users content consumption history and other factors that further adds to the ideological bias in the content consumed by the user.

This research focuses on using data mining techniques to study information filter bubble on twitter. We aim to uncover an ideological latent space which would help us understand the ideological stances of twitter users and their association with news sources. Our goal is to use this ideological latent space to collectively address two problems that are central to tackling the problem of information filter bubble (i) how can we make the users aware of their own information bubble (ii) how can we diversify the content that they consume. To this end we start with the following preliminary research questions

- RQ1: First, is there any ideological polarization of users and content on Twitter?
- RQ2: Can we use Twitter information (such as tweet, re-tweet and follows) to estimate ideological leaning of users and content?
- RQ3: How can we this estimated Ideological Leaning of users and content on Twitter to diffuse the information filter bubbles?

1.2 Literature

In the recent times there has been an increased interest in tackling the problem of information filter bubbles. Wall street journal developed a user interface [2] to raise awareness of the difference in view points by presenting side-by-side articles from *blue feed* and *red feed*. Multiple browser extensions have been created to remind user of the imbalance of reading [28]. *Escape your bubble* [3] is a browser extent which adds manually curated event from the other side to a users news feed.

There has also been some attention from the research community to reduce the information filter bubble [34] [33]. Several studies have been carried out to show the existence of filter bubbles online [6] [10] and the negative effects caused by filter bubbles [21] [35]. A large body of research work has studied and quantified the filter bubble [10] [27] [15] [5] [39].

Recently there have been a few attempt to reduce this polarization, specially in the direction of exposing users to opposing viewpoints. Much of the work in this area focuses on "how" to recommend content from the opposite side by designing interfaces and user surveys [19] [28] [30] [16]. Garimella et al. [17] propose to reduce the polarization by recommending content from opposite side to users. In their work they focus on the problem of "who" to recommend by modeling user's "acceptance" probability using a random walk based approach. We observe that "what to recommend" is largely an open problem. We believe it would be interesting to tie the problem of "what to recommend" and "who to recommend" and holistically address the problem in an algorithmic way.

Further, most of the work in the literature focuses on only two-sides of a controversy. They aim to separate the user into two ideological clusters - *liberal* and *conservative* and focus on *how* to recommend content to the *opposite* side. However, learning a binary stance, is hugely undermining the problem. In a real world, users do not have binary ideologies, instead each user has a degree to which he confirms to a certain ideology. Knowing this ideological position is critical to answer two important questions (i) "who"(user) to recommend (ii) "what"(content) to recommend. For instance, one can imagine that, it would be easier for an extreme user to look at neutral content and perhaps even absorb some of the information. On the other hand, exposing an extreme user to ideologically extreme content from the other side can further increase the polarization [19]. Likewise, identifying a neutral user can be fruitful in connecting the two ideologically separated sides. For example, these neutral users, who are part of the conversation in both the ideological clusters can be very helpful in relaying the information between the two clusters and bridging the information gap. Further, studies have shown that *neutral* users are easier to nudge and it is more likely that a content from the opposite side would be consumed by a neutral user [25]. Hence it is not sufficient to separate the users into ideological clusters. In order to make any meaningful recommendations we need to estimate the position of a user on a continuous ideological scale so that appropriate recommendations can be made to diversity the content.

1.3 Research questions

Having considered the open problems in the literature we revisit the research questions that we initially started with. $RQ1$ has been extensively studied and addressed in the literature. However $RQ2$ and $RQ3$ seem very promising especially in the light of (i) the value in learning the ideological position of users and sources on a continuous scale (ii) the possibility to tie the problem of "what" and "who" to recommend and learn the two simultaneously. Interestingly, if we consider "what" as the content to be recommended and "who" as the user to recommend to, we can model this as a problem to learn ideological stance of both users and sources simultaneously. With the new found motivation, we reformulate our research questions as follows:

- RQ2: Can we use Twitter information (such as tweet, re-tweet and follows) to estimate a continuous ideological score for both users and content in a unified formulation?
- RQ3: How can we this estimated ideological leaning of users and content to diversify the content consumed by a user by tying together "who" and "what" to recommend?

1.4 Contributions

To answer the first question (RQ2), we use public twitter data to uncover Twitter users' and popular news media channels' ideology. We collect 19 million tweets over a span of 5 years from 2011 to 2016. For our study, we filter tweets from 559 popular news media outlets about 3 controversial topics – Obamacare, Gun Control and Abortion. We represent the tweets as a matrix where one dimension is the user and the other dimension is the source (news media channel) of the tweet. The other input to our algorithm is the adjacency matrix on the social graph of users. Given these two inputs our matrix factorization based algorithm uncovers the underlying multidimensional ideological latent space in Twitter of both users and sources at the same time. We use this ideological latent space to compute the ideology scores for users and sources.

In order to address the second question (RQ3), we present two case studies. First, we call for raising the awareness of social network users by providing visual, explanatory and exploratory evidence of their own information filter bubble. We provide an interactive platform to the users to explore their own content consumption bias (user as well as algorithmic). Thus, we lower the barrier for understanding an information filter bubble for users without any domain knowledge. Second, we present preliminary work to increase the diversity and transparency of user's content consumption by discussing two possible solutions of (i) incorporating the learnt latent space of user and source in the recommendation system itself and (ii) designing exploratory and interactive interfaces for the users to explore the content and self-adjust diversity of their content consumption.

1.5 Outline

The rest of the thesis is organized in the following chapters.

Chapter 2 is an introductory chapter, which presents some of the terms and methods which are central to the thesis. This chapter introduces the reader to Twitter terminology, social science terms used in the thesis and presents an overview of non-negative matrix factorization.

In Chapter 3 we introduce the reader to definitions and notations, and formulate the problem from a mathematical standpoint.

In Chapter 4 we build on the problem formulation described in chapter 3 and present an end-to-end framework to compute ideological leaning for users and sources.

Chapter 5 details experimental setup, baseline algorithms, evaluation measures and analysis of results.

In Chapter 6 we present two case studies and prototypes on how the ideological scores learnt in chapter 4 can be used to reduce the information filter bubble.

Finally in Chapter 7 the topics covered in introduction of the thesis are summarized and main conclusions are discussed.

2 Background

This chapter is a review of the data mining, machine learning and social science methods and terms most central to this thesis.

2.1 Terms and Definitions

In this section we list and define commonly used domain specific terminology.

2.1.1 Information Filter Bubble

The term filter bubble was coined by Pariser [29]. It is defined as the phenomena where users get less exposure to conflicting viewpoints and are intellectually isolated in their own informational bubble. As mentioned earlier in Chapter 1, social networks, personalization algorithms such as recommendation systems, and search engines directly or indirectly impose many layers of filters on the content that a user can see on the Internet. These filters, when combined with the users' information bias and preferences, leads to the scenario where users are effectively separated from the information that disagrees with their viewpoints, consequently isolating them in their own ideological filter bubbles. In the recent studies many terms have been used synonymous to the term "filter bubble", for example, "echo chambers", "ideological silos". Other terms such as "ideological polarization", "ideological segregation" and "controversy" are also closely related to *filter bubble*, in the sense that they can be seen as the consequences of the *information filter bubbles*.

2.1.2 Twitter Terminology

Twitter is a social media platform where the core of the platform is an element called *Tweet*. A Tweet is a 140 character long public post which can be seen by any one on Twitter. A tweet typically contains some text, and in some cases a link to an external URL (content). Users on the platform have three types of interaction options to connect with other users: follow, retweet and user mention.

- Every Tweet is open for all Twitter users to read, comment and to re-post, such a rep-post is call a *retweet*. Typically a retweet represents endorsement of the content mentioned in the tweet. However, in some cases it could represent a different justification, for instance, someone making fun of the post or criticizing the post.

- *Follows* is a directed relationship where a user subscribes to view content (tweets) posted by another user. It is a popular notion that the follows relationship between users represents ideological similarity. However, in some cases one could follow a person from the opposite camp so as to be informed about what is happening (e.g., a lot of Democrats follow Trump)

- In a *user mention*, a Twitter-user tags another Twitter-user usually to discuss something related to the said user or to gain their attention.

Since it is challenging to determine if a *user mention* indicates an *endorsement* or *contention*, in our thesis, we do not use user mention to derive ideological relationship between users.

In this thesis we focus on two kind of data types in Twitter : (i) Twitter-users and (ii) Twitter content.

Twitter users are the set of users who use the social platform Twitter. For the purpose of this thesis we call the set of Twitter users collected in the dataset as *users*.

The primary content on Twitter is a *tweet*, as mentioned earlier, many of the tweets contain an external URL. We refer to the set of extent URLs tweeted by *users* as *content*, and the set of news media channels which are the author of the *content* as *source*. That is, a *source* is the hostname extracted from the content URLs.

2.2 Overview of NMF

Non-negative Matrix Factorization (NMF) [23] has attracted a lot of attention in the last decade in machine learning and data mining, especially in the context of clustering due to its ease of interpretability. In this section we provide a comprehensive overview of NMF methods for clustering.

2.2.1 NMF for Clustering

Given an input matrix of the form $X \in \mathbb{R}^{n \times m}$, we seek to factorize X into two latent factors,

$$X \approx UV^T \tag{1}$$

where $U \in \mathbb{R}^{n \times k}$ and $V \in \mathbb{R}^{m \times k}$ are non-negative matrices, and k is the number of clusters defined in the clustering.

When NMF is applied for clustering, each row in X is treated as a data point in an N dimensional space. In such a case, the factors U and V have the following interpretation:

- U_{ij} corresponds to the degree to which row i belongs to cluster j
- V_{ij} corresponds to the degree to which column i is associated with cluster j

with appropriate normalizations, V_{ij} is proportional to the posterior probability of cluster j given column i.

Our goal is to find U and V to minimize J. The most common cost function used for the minimization of the objective function is the **sum of square errors**. In this thesis, we assume the matrix norm to be the Frobenius norm.

$$J = \min_{U \geq 0, V \geq 0} J\|X - UV^T\|_F^2 \tag{2}$$

2.2.2 Equivalence between NMF and K-Means

An equivalence between NMF *sum of square error* optimization with orthogonality constraints and K-Means was proven by Ding et al. [11] where they show that

$$J = \min_{U \geq 0, V \geq 0} J\|X - UV^T\|_F^2 \text{ s.t. } VV^T = I \tag{3}$$

is equivalent to K means clustering, U has the meaning of cluster centroids and V is equivalent to cluster indicator matrix.

2.2.3 Co-clustering

NMF seeks one-sided clustering of either rows or columns of the input matrix. However, for many real-world applications we seek to infer both row and column clusterings simultaneously, as well as the association between the row and column clusters. For instance, a typical document clustering problem involves two different data types *documents* and *terms*. In such a co-clustering problem we seek to co-cluster both documents and terms simultaneously by leveraging the dual relationship between the two data types.

2.2.4 NMF with Tri-Factorization (NMTF)

NMTF [12] provides a good foundation to perform co-clustering. In NMTF our goal is to simultaneously cluster rows and columns of the input matrix X. NMTF seeks a 3-factor decomposition of matrices

$$X \approx UHV^T \tag{4}$$

where $U \in \mathbb{R}^{n\times\ell}$, $H \in \mathbb{R}^{\ell\times k}$ and $V \in \mathbb{R}^{m\times k}$ are non-negative matrices l is the number of row clusters and k is the number of column clusters defined in the clustering.

More precisely, we solve the optimization problem with dual orthogonality constraints

$$J = \min_{U\geq 0, H\geq 0, V\geq 0} \|X - UHV^T\|_F^2 \text{ s.t. } UU^T = I, VV^T = I \tag{5}$$

where U is the row coefficient matrix which gives row clusters and V is the column coefficient matrix which gives column clusters. H provides the association between row and column clusters. It provides additional degree of freedom and allows accurate clustering of both rows and columns simultaneously.

2.2.5 Symmetric Matrix Factorization

Symmetric matrix factorization is a special case of non-negative matrix factorization in that the input to the factorization is a matrix of pairwise similarities: $X = X^T$. As a consequence the row and column latent factors are the same $U = V$. The symmetric NMF optimization is:

$$\min_{U\geq 0, H\geq 0, V\geq 0} \|X - UHU^T\|_F^2, \text{ s.t. } UU^T = I \tag{6}$$

Given the input X as the adjacency matrix of a Graph, it is well known that the matrix factorization in Equation (6) is equivalent to Normalized Cut spectral clustering [11] [37].

2.2.6 Graph Regularization

Cai et al. [9] proposed a graph based approach on NMF called Graph regularized Non-negative Matrix Factorization (GNMF) that considers geometric structure in the data. In their paper, they encode the geometric information in the form of an affinity graph and seek a matrix factorization, which respects this graph structure. That is, if two points are connected in the affinity graph these two data points should be sufficiently close to each other in the matrix factors. They refer to this assumption as *manifold assumption* and formulate it as follows:

$$\frac{1}{2}\sum_{i,j}\|u_i - u_j\|^2 (W_v)_{ij} = \mathrm{tr}(U^T L^u U) \tag{7}$$

where W_u is the affinity graph defined for the nodes in G, $L_u = D_u - W_u$ is the graph Laplacian on the affinity graph W_u, D_u is a diagonal degree matrix of W_u such that $D_u = \sum_{j=1}^{N}(W_u)_{ij}$.

The new matrix factorization (GNMF) incorporates the graph structure into its objective function as follows:

$$J = \min_{U\geq 0, V\geq 0}\|X - UV^T\|_F^2 + \alpha \cdot \mathrm{tr}(U^T L^u U) \tag{8}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, L_u is the graph laplacian on any adjacency matrix consisting of rows of X, α is a regularization parameter and the term $\mathrm{tr}(U^T L^u U)$ is called the graph regularization constraint, which captures the geometric structure in the input data.

Gu and Zhou [20] proposed a Dual manifold regularized co-clustering method (DMCC) based on the duality between rows and columns of the co-clustering matrix, which imposes manifold regularization on both rows and columns of input matrix, i.e., it considers geometric structure on both the row and column manifolds. For an input matrix X, DMCC has the following objective function

$$J = \min_{U\geq 0, H\geq 0, V\geq 0}\|X - UHV^T\|_F^2 + \alpha \cdot \mathrm{tr}(U^T L^u U) + \beta \cdot \mathrm{tr}(V^T L^v V) \tag{9}$$

where tr(.) denotes the trace of a matrix, H reflects the association between rows and columns of X, α and β are regularization parameters, L_u is the graph laplacian on any adjacency matrix consisting of rows of X and L_v is the graph laplacian on any adjacency matrix consisting of columns of X.

3 Problem Formulation

3.1 Motivation

The main premise of this thesis is that the tweet and retweet behavior of a user on Twitter is consistent with his *liberal-conservative* ideological leaning.

A simple approach to identify ideological leaning of a user is to treat user interactions/relationships as a social graph or adjacency matrix. One could then apply any of the graph partitioning algorithms to detect communities of users. However, relying meagerly on user relations has two major problems: (i) In general, in any of the real world social networks such as twitter, the number of user connections is much lower than the number of total users. As a consequence, the underlying graph is highly sparse. (ii) Clustering which depends only on links between users ignores one fundamental assumption of ideology clustering, i.e., users who are similar to each other should be in the same cluster. Consider two users u_1 and u_2 who have the same ideology and share similar content. However, they do not know each other and have no social relationship. We would still like the users u_1 and u_2 to be in the same cluster even if they do not have any relations between them.

For community detection of sources, one simple approach is to consider sources as individual entities and perform NLP-based semantic and sentiment analysis techniques on the content information to identify source clustering. However, this approach is complex and prone to errors due to the noise in the data. Another approach is to construct a link graph between sources based on their similarity (e.g., the number of common users/shares/clicks etc). However, this approach still ignores the rich user link information.

Therefore, combining the two types of input to collectively learn the communities of sources and users may be a better strategy. Furthermore, all these techniques only have one sided clustering of either user or content, with no association between them.

Thus, our motivation in combining sources of information is many-fold: (i) It allows us to learn collectively from two different types of input. Clearly, clustering with increased information will result in better clustering performance. (ii) It allows simultaneous learning of the two types of clustering in a unified seamless approach. (iii) Furthermore, a shared latent space keeps the dual interconnected relationship between users and sources intact. That means, not only do we want to separate users and sources into ideological clusters, we would also like to know the relationship between the two clusterings. Simultaneous clustering with shared latent variables provides an explicit and compact representation of both ideological clusterings of user and sources as well as the relationship between them.

3.2 Summary

Given a Twitter user's social link information (re-tweet and follows) and content creation/consumption (tweet and re-tweet) information, we can construct a social matrix (user $\times$ user) and an associated bipartite content matrix (user $\times$ content). We aim to learn a common ideological latent space shared by both users and content, such that knowledge of user interactions in the link matrix can be transferred to content clustering, and knowledge of user-content interaction from content matrix can be transferred to user clustering.

Many techniques have been proposed to learn the latent space in input data. Non-negative Matrix Factorization (NMF) is one of the most popular techniques due to its ease of interpretability and performance. In this work, we propose a joint matrix factorization formulation, which exploits the duality between user and source clustering. We learn a shared ideological latent space using iterative multiplicative update rules to solve the optimization problem.

In the proposed technique, both the user's social link information as well as his/her content information are seamlessly combined in a unified approach with a set of shared *latent factors*. Our model contains two latent factors.

The first component (which we later refer as U) captures the user information. That means, information about users is decomposed into latent user $\times$ ideology factors where the rows of the matrix represent the degree to which each row ID (user) belongs to the said column (ideology) cluster.

The second component of the factorization (which we later refer as V) represents the source ideology factors. Sources are decomposed into latent ideology $\times$ source factors. The column vectors represent the degree to which the said column belongs to the row cluster. The bridge latent factor (H) of size $k \times k$, where k is the number of latent dimensions, captures the association between the corresponding row and column clusters. For instance, if the H matrix is close to an identity matrix, the diagonal of the matrix suggests that there exists a pairwise association between corresponding row and column clusters.

We connect the two components (U and V) through a shared set of latent factors by modeling the problem as a joint factorization problem (on *social link graph* and *user-content information*) and tying the latent factors in the two matrix decompositions. As a consequence, the optimization problem simultaneously and seamlessly searches for a shared latent space for both the components that best explains both link and content structure.

3.3 Definitions and Notations

Suppose we have a social graph of twitter users, $G = (\mathcal{V}, \mathcal{E}, w)$, where $\mathcal{V} = \{x_i\}_{i=1}^N$ represents the set of users, $\mathcal{E}$ is the set of edges representing social interaction between the users and w is the set of edge weights. The adjacency matrix of links between users associated to the graph G is denoted by $A \in \mathbb{R}^{n \times n}$. For a pair of nodes x_i and x_j, $A(i, j) = w$ if there is an edge between x_i and x_j, and $A(i, j) = 0$ otherwise. Note that in this definition, A may be either symmetric of non-symmetric based on the underlying graph G. Each node $x_i \in G$ has a set of associated content features denoted as $C_i = \{C_{ij}\}_{j=1}^m$, where the feature set represents the content on Twitter. Thus, the bipartite content matrix associated to the graph G is denoted as $C \in \mathbb{R}^{n \times m}$.

We derive the edges in G from social links between users based on their re-tweet and follows network. Specifically, we experiment with two variants: For a pair of users u_i and u_j, there is an edge $e = (u_i, u_j, w)$ (i) if u_i and u_j *follow* a common set

Table 1: Notations used in this thesis and their corresponding explanations and dimensions

Notations	Details	Dimensions
G	endorsement graph	-
$\mathcal{V}$	set of users	-
$\mathcal{E}$	set of directed edges	-
w	set of edge weights	-
n	number of users	-
m	number of sources	-
k	number of clusters	-
A	user-user social matrix associated to graph G	$n \times n$
C	user-source contant matrix	$n \times m$
U	user-cluster matrix	$n \times k$
V	source-cluster matrix	$m \times k$
H_1	cluster association matrix for user	$k \times k$
H_2	cluster association matrix for source	$k \times k$
L^u	graph laplacian matrix for user	$n \times n$
L^v	graph laplacian matrix for source	$m \times m$
α, β	parameters to control influence of graph regularization	-
$tr(\cdot)$	trace of a matrix	-
$\|\cdot\|_F$	frobenius norm	-

of users, and w is the size of intersection of *follows*, or (ii) if u_i retweets u_j, then w is the number of retweets. We derive the content features in C associated to the graph G based on users' tweets. Specifically, we experiment with two variants of feature sets: (i) extracted urls from the tweets and (ii) hostnames of urls extracted from the tweets. Therefore, the content is aggregated by their source of authorship (i.e., various news media channels).

3.4 Problem Description

Given a social matrix A and a content matrix C, we would like to consider the following constraints in order to determine the user and source ideology clusters:

Partitioning constraints: We want to take into account the network structure in the link matrix A as well as the user-content information in the bi-partite content matrix C in order to partition the users and content into ideology clusters such that the following constraints hold true:

1. Users in a user cluster interact with each other more often than outside the cluster.

2. Users in a user cluster consume content which is more similar to the content consumed by within cluster users than with users outside the cluster.

Co-partitioning constraints: We want to identify the association between the resulting user and source partitions so that

3. users in a user partition share more articles from their corresponding content partition than from other content partitions.
4. content in a content partition is shared by more users from their corresponding user partition than from other user partitions.

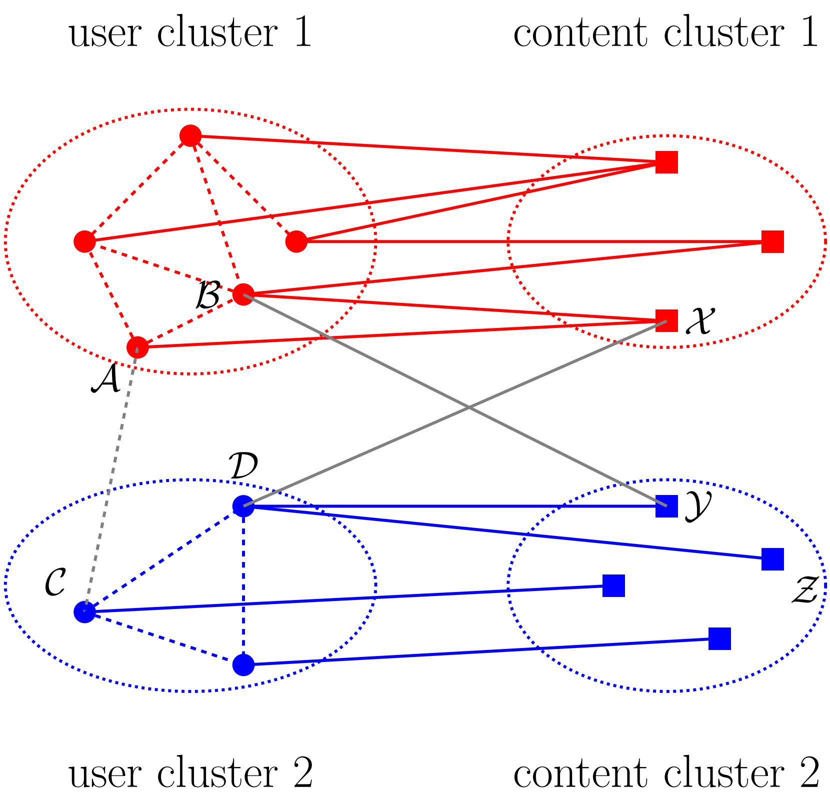


Figure 1: Pictorial representation of the partitioning and co-partitioning constraints. As per the constraints discussed in Section 3.4, users $\mathcal{A}$ and $\mathcal{B}$ belong to the same user cluster and their shared content $\mathcal{X}$ is in the corresponding content cluster. Users $\mathcal{C}$ and $\mathcal{D}$ belong to the same user cluster and content $\mathcal{Y}$ and $\mathcal{Z}$ in the corresponding content cluster.

3.5 Problem Formulation

We derive our problem formulation for learning ideological space in multiple steps: (i) We present the basic non-negative matrix factorization model for co-clustering of users and content. (ii) We derive the latent factorization model for combining Link and Content as well as for learning a shared latent space. (iii) We summarize the co-partitioning criteria described in Section 3.4 as geometric manifold constraints and encapsulate them in the model by adding graph regularization constraints.

(i) Non-negative Matrix Factorization for Co-Clustering: For a given input data matrix X, the bi-orthogonal non-negative 3-factor decomposition (ONMTF) introduced by Ding et al. [12] provides a good framework for simultaneously clustering the rows and the columns of the X. The rows of the latent factor matrix U provide clustering of users while columns of latent factor matrix V provide article clustering. The cluster association matrix H absorbs the difference in scales of X, U and V and provides additional degrees of freedom. The presence of H ensures that the low-rank matrix representation remains accurate, while U gives row clusters and V gives column clusters.

$$X \approxeq UHV^T \tag{10}$$

With the notation and formulation used above, the co-partition detection problem in this thesis is formally defined as follows

Problem 1 *(ONMTF [Ding et al. [12]]) Given a bi-partite user-source input matrix C of dimension $n \times m$. Find non-negative matrices U and V of dimensions $n \times k$ and $m \times k$ respectively such that we minimize*

$$J_1 = \min_{U \geq 0, H_2 \geq 0, V \geq 0} \|C - UH_2V^T\|_F^2 \tag{11}$$

$$s.t.\ UU^T = I, VV^T = I, \tag{12}$$

where $\|.\|_F$ denotes the Frobenius norm.

An important special case is when the rows and columns of input matrix are indexed by the same set of objects. In our case, the links in the rows and columns of the user-user link matrix A belong to the same objects(i.e., users). Hence, we can re-write $U = V = U$ in Equation (11)

$$J_2 = \min_{U \geq 0, H_1 \geq 0} \|A - UH_1U^T\|_F^2 \tag{13}$$

$$\text{s.t. } UU^T = I \tag{14}$$

Note that H_1 is not necessarily symmetric. That means, UH_1U^T can produce non-symmetric matrix, which is the case of A (user-user social link matrix) in our case. Furthermore, an important advantage of this form of tri-factorization is that it captures link transitivity [42]. Consider a transitive link $u_i \rightarrow u_s \rightarrow u_j$ between u_i and u_j, where the user u_i is linked to u_j via u_s. A non-symmetric factorization of the form $A \approxeq ZU^T$ treats the values in input matrix A as links from set of users to a different set of objects, let it be denoted by $\mathcal{O} = \{o_i\}$. Hence it would split the link path $(u_i \rightarrow u_s \rightarrow u_j)$ into two parts $u_i \rightarrow o_s$ and $u_s \rightarrow o_j$ which is a misinterpretation of the original link path. Whereas, $A \approxeq UH_1U^T$ considers the links to be amongst the same set of objects. Hence, the transitive link $u_i \rightarrow u_s \rightarrow u_j$ is correctly captured by latent factors in U.

(ii) Combining Link and Content: The basic assumption in combining multiple input datatypes is that the input matrices may share common knowledge structures which can be captured by a common shared latent space. We aim to learn a common ideological latent space shared by both users and content, so that knowledge of user interactions in the link matrix can be transferred to content clustering, and knowledge of user-content interaction from content matrix can be transferred to user clustering. Zhu et al. [42] proposed an algorithm to classify web-pages by exploiting both content and link information of the web-pages. They carry out a joint-factorization on both the link adjacency matrix and the document-term matrix. By mapping both types of information onto a common low dimensional latent space, they derive a low-dimensional latent representation for web-pages, without explicitly separating them as content and link factors. Inspired by this idea, we aim to combine the information from both the data-types, i.e., link information (A) and content information (C).

With the motivation defined above, the problem of combining link and content information in A and C, as well as the simultaneous partition detection of user and source can be written as a natural extension to Problem 1. We formally define the problem as follows:

Problem 2 *Given a user-user social matrix A of dimension $n \times n$, and a bipartite user-source content matrix of dimensions $n \times m$, find non-negative matrices U and V of dimensions $n \times k$ and $m \times k$ respectively such that we minimize:*

$$J_3 = \min_{U\geq 0, H_1\geq 0, H_2\geq 0, V\geq 0} \|A - UH_1U^T\|_F^2 + \|C - UH_2V^T\|_F^2 \tag{15}$$

$$s.t.\ UU^T = I, VV^T = I \tag{16}$$

where $tr(\cdot)$ denotes the trace of a matrix, U and V represent the learnt factors for users and content respectively in the ideological latent space.

In Equation (15) we carry out a joint factorization on both user-user link matrix (A) and user-article content bi-partite matrix (C). In order to learn a common ideological latent space for both users and articles, we tie the latent factor U in Equation (13) with the latent factor U in Equation (11).

(iii) Graph Regularizations: The formulation for clustering in ONMTF fails to consider the geometric structure in the data, which is essential for clustering data on manifold. To address this problem Cai et al. [9] introduced graph regularized NMF based on the *manifold assumption* that, if two data points x_i,x_j are close in the intrinsic geometry of the data distribution in X, then the representations of this two points u_i and u_j in the new basis U are also close to each other. This is formulated as follows,

$$\frac{1}{2}\sum_{i,j}\|u_i - u_j\|^2(W_v)_{ij} = \mathrm{tr}(U^T L^u U) \tag{17}$$

where W_u is the affinity graph defined for the nodes in G, $L_u = D_u - W_u$ is the graph Laplacian on the affinity graph W_u, D_u is a diagonal degree matrix of W_u such that $D_u = \sum_{j=1}^{N}(W_u)_{ij}$.

Gu and Zhou [20] proposed a Dual manifold regularized co-clustering method based on the duality between rows and columns of the co-clustering matrix which imposes manifold regularization on both rows and columns of input matrix, i.e., for an input matrix X

$$J = \min_{U \geq 0, H \geq 0, V \geq 0} \|X - UHV^T\|_F^2 + \alpha \cdot \mathrm{tr}(U^T L^u U) + \beta \cdot \mathrm{tr}(V^T L^v V) \tag{18}$$

where H reflects the association between rows and columns of X. L_u is the graph laplacian on any adjacency matrix consisting of rows of X and L_v is the graph laplacian on any adjacency matrix consisting of columns of X.

We extend the Problem 2 to include dual graph regularization constraints on users and sources. At the end, the co-partitioning problem of users and sources on an ideological plane is defined as the joint non-negative matrix tri-factorization with bi-orthogonality constraints and dual graph regularization constraints. We formally define the final problem as follows:

Problem 3 *Given a user-user social matrix A of dimension $n \times n$, and a bipartite user-source content matrix of dimensions $n \times m$, find non-negative matrices U and V of dimensions $n \times k$ and $m \times k$ respectively such that we minimize:*

$$J_4 = \min_{U \geq 0, H1 \geq 0, H3 \geq 0, V \geq 0} \|A - UH_1U^T\|_F^2 + \|C - UH_3V^T\|_F^2 \tag{19}$$

$$+ \alpha \cdot tr(U^T L^u U) + \beta \cdot tr(V^T L^v V) \tag{20}$$

$$s.t.\ UU^T = I, VV^T = I \tag{21}$$

where A and C are input matrices, U, V, $H_1$and H_3 are non-negative latent factor matrices, α and β are parameters to control the influence of user and source affinity graphs L_u and L_v in the joint factorization.

This formulation factorizes A and C jointly based on the dual manifold assumption i.e., both users and content share the same latent space and the cluster labels of users are smooth with respect to the content manifold, while the cluster labels of content are smooth with respect to user manifolds. In order to apply manifold constraints, first, we construct affinity graphs for users W_u (and content W_v). While there are many ways to construct such an affinity graph, in our experiments we construct W_u (and W_v) as cosine similarity matrix on row vectors (and column vectors) of the content matrix C. We then construct their corresponding graph Laplacian matrices L_u (and L_v), such that they capture the geometric structure of data manifold (and feature manifold). Next, we encode this geometrical information in the joint factorization model in Equation (15) in the form of graph regularization constraints. We then seek a matrix factorization that respects this graph structure.

3.6 Optimization Problem

In this section we solve the optimization problem and derive multiplicative update rules.

3.6.1 Optimization problem

Following the standard theory of constrained optimization, we solve the following optimization problem in Equation (3) by introducing Lagrangian multipliers λ (a symmetric matrix of size $K \times K$) and minimizing the Lagrangian function

$$L = \min_{U\geq 0, H1\geq 0, H3\geq 0, V\geq 0} \|A - UH_1U^T\|_F^2 + \|C - UH_3V^T\|_F^2 \tag{22}$$

$$+ \alpha \cdot \mathrm{tr}(U^T L^u U) + \beta \cdot \mathrm{tr}(V^T L^d V) \tag{23}$$

$$+ tr[\lambda(UU^T - I)] + tr[\lambda(VV^T - I)] \tag{24}$$

3.6.2 Derivation of update rules

We can compute the gradient of L with respect to U,V,H_1 and H_3. The optimal solution to the optimization in Equation (22) can be achieved using an iterative update algorithm proposed by Ding et al. [12]. The update rules are as follows:

$$U \leftarrow U \cdot \sqrt{\frac{AUH_1^T + CVH_3^T + \alpha \cdot S_u U}{UH_1U^TUH_1^T + UH_3V^TVH_3^T + \alpha D_u U + U\lambda_u}} \tag{25}$$

$$V \leftarrow V \cdot \sqrt{\frac{C^TUH_3 + \beta \cdot S_d V}{\beta \cdot D_d V + VH_3U^TUH_3 + V\lambda_v}} \tag{26}$$

$$H_1 \leftarrow H_1 \cdot \sqrt{\frac{U^TAU}{U^TUH_1U^TU}} \tag{27}$$

$$H_3 \leftarrow H_3 \cdot \sqrt{\frac{U^TCV}{U^TUH_3V^TV}} \tag{28}$$

where

$$\lambda_u = U^TAUH_1^T + U^TCVH_3^T - \alpha U^T \cdot L_u U - H_1U^TUH_1^T - H_3V^TVH_3^T$$
$$\lambda_v = V^TC^TUH_3 - \beta V^T \cdot L_d V - H_3U^TUH_3$$

4 Methodology

4.1 Summary

In this section, we present the proposed NMTF based end-to-end framework for estimating Ideological leaning for Twitter users and Media channels. The core of the proposed method is the NMTF based optimization problem to learn the latent factors discussed in Chapter 3. Next, we utilize the probabilistic model of NMF factorizations discussed in Yoo and Choi [40] to derive a probabilistic interpretation of our latent factors. Finally, we present how these latent factors can be used to infer ideological leaning of Twitter users and Media Channels.

The most interesting contribution of this section is how a continuous score of ideological leaning of Twitter users and Media Channels can be inferred from open Twitter data. For the sake of completeness, we also show how a hard cross-ideological separation (hard clustering) can be derived from the latent factors.

4.2 Learning Ideological Latent Space

Figure 2 presents a logical flowchart of all the steps involved in learning user and source ideology scores. First, we combine user's social link information and their content consumption information, and learn a shared ideological latent space by applying the NMTF based latent space model discussed in Chapter 3. Next, we normalize the factors to derive probabilistic interpretation of the factors. We then transform these latent factors to compute ideology and popularity scores. The details about these components are presented in the following sections.

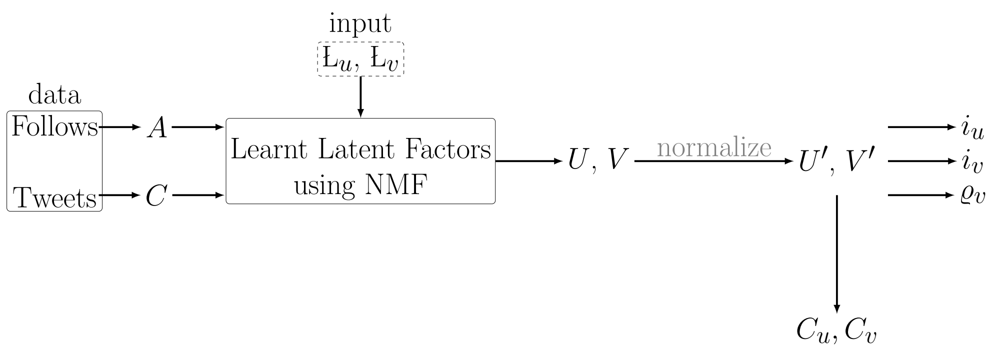
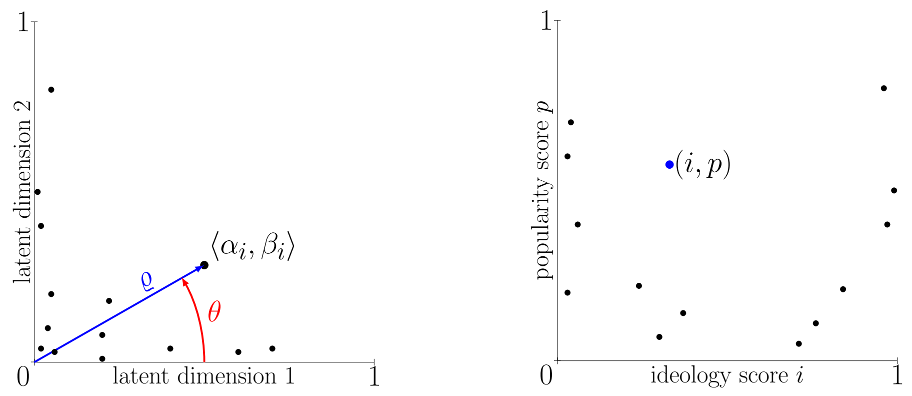


Figure 2: Flowchart of Proposed Methodology

4.3 Probabilistic Interpretation of Latent Factors

The latent factors (U,V) learnt from this factorization have a probabilistic interpretation as follows:

– U_{ij} corresponds to the degree to which user u_i belongs to the user-cluster $\{c_u\}_j$.



(a) latent space **(b)** transformed latent space

Figure 3: Projection of data points

– V_{ij} corresponds to the degree content v_i is associated with content-cluster $\{c_v\}_j$.

With appropriate normalization as proposed in the literature in (Yoo and Choi [40]) & (Li and Ding [24]), U_{ij} is proportional to a posterior probability of user-cluster $\{c_u\}_j$ given user u_i, and V_{ij} is proportional to a posterior probability of content-cluster $\{c_v\}_j$ given content v_i. In our case, the graph regularization constraints in our factorization separate the clusters by *liberal-conservative* ideological similarity. Since we have set the number of latent dimensions to 2, the factors U_{ij} and V_{ij} are proportional to the posterior probability of belonging to *liberal* and *conservative* clusters.

4.4 Projecting Latent Factors

Say, α_u and β_u represents the two column vectors of U and α_v and β_v represents the two column vectors of V. We plot the data in the subspace of the two column vectors α and β such that each data point $(\alpha_{ui}, \beta_{ui})$ represents a user in the *user ideological latent space*, each data point $(\alpha_{vi}, \beta_{vi})$ represents a content in the *content ideological latent space* and each axis corresponds to a cluster. This model has a simple visual interpretation:

(i) All the data points belonging to the same cluster are located close to the axis. It is clear that the cluster label for a data point can be determined by finding the axis with which the data point has the largest projection. Extending the idea, if we consider each data point as the posterior probability vector $\langle\alpha_{vi}, \beta_{vi}\rangle$, (i) The data point belongs to the axis with which it has the least angle (ii) say x-axis (0°) corresponds to *liberal cluster* and y-axis (90°) corresponds to *conservative cluster*. We could calculate a normalized ideology score in range 0 (liberal) - 1 (conservative) as the normalized angle θ of the posterior probability vector $\langle\alpha_{vi}, \beta_{vi}\rangle$

$$i = \frac{\arctan(\frac{\beta_i}{\alpha_i})}{\pi/2} \tag{29}$$

(ii) The position of the data point on the axis is proportionate to the strength of its value in the latent matrix. Since the latent matrix is meagerly a factorization of the input matrix, this value is directly correlated to the value in the input matrix, i.e., the user's engagement on Twitter (retweets and follows) and the sources popularity on Twitter (number of tweets containing the source). Therefore, the magnitude ϱ of the posterior probability vector $\langle \alpha_{vi}, \beta_{vi} \rangle$ is equivalent to the popularity (engagement on Twitter). Thus, the normalized popularity/engagement score can be derived as

$$\varrho = \sqrt{{\alpha_i}^2 + {\beta_i}^2} \tag{30}$$

Figure 14 visualizes the data points projected in the original latent space and their transformation to the corresponding ideology/popularity co-ordinate space. It is interesting to note that, in our quest for uncovering "ideology" of users and source, "popularity" was a dimension we discovered serendipitously. However, it is easy to understand the reasoning behind it. The value C_{ij} in the input matrix C captures the number of tweets tweeted by i-th user that were authored by j-th source. A popular (highly tweeted) source would thus have a dense column with large magnitude in the input matrix which would be reflected by larger magnitude in the corresponding latent factorization. In order to validate if the dimension indeed captures the popularity of a source, we compute ground of popularity score for each source using the number of tweets that contain the source as a proxy for its popularity. We observe that there is a pearson correlation coefficient of 0.9 between the estimated popularity score from the latent space and the computed ground truth.(details follow in Section 5)

While "popularity" is an accidentally discovered dimension, it is an extremely interesting dimension to address the problem of reducing filter bubbles. Garimella et al. [17] suggest that highest reduction in user-polarization score is achieved by connecting a user with authoritative source with the opposing view. That is, maximum reduction in user-polarization happens when a user is recommended "popular" content from the opposing viewpoint. The intuition behind the idea is that this way the user can see the "popular" notion on the other side, and presumably the "popular" content is usually of good quality. We do not investigate further in this direction as it is out of the scope of this thesis. However, we use the dimension "popularity" in the case studies presented in Section 6.

5 Experiments

In this section we present a set of experiments on real world datasets to validate the effectiveness of our NMF based ideology learning technique. In our experimental evaluation our focus is to evaluate the computed ideological scores as well the ideological cluster separation for users and sources.

5.1 Dataset collection and processing

The dataset that we use is collected using Twitter's streaming API (random 1% sample) from 2011 to 2016. We selected three popular controversial/polarized topics on Twitter which discussed by a large number of Twitter users with both opposing sides of ideology: "gun control", "abortion" and "obamacare". We collect all tweets (and their corresponding users) related to these topics by filtering based on topic-related hashtags and related keywords [26]:

– gun control: gun control, gun right, pro gun, anti gun, gun free, gun law, gun safety, gun violence

– abortion: abortion, prolife, prochoice, anti-abortion, proabortion, planned parenthood

– obamacare: obamacare, #aca

Due to the 1% random sample given by Twitter API our dataset for users was quite sparse. Majority of the users tweet once or twice. In order to negate this problem, and have a meaningful dataset we filtered the set of users for which we had decent amount of twitter activity. To this end, we obtained all users who had tweeted on all the three topics. This gave us a set of 6391 users. We then collected all the tweets of these 6391 users. Note that due to twitter API restrictions we could collect up to 3200 tweets for each user.

Since our goal is to be able to identify ideology of both the user as well as the source, we filter the tweet set to contain only tweets from well known news media channels in the US region. For this purpose we aggregate a set of 559 news domains (and their shortened url versions) obtained from previous work in the literature [7][14][18]. Fortunately, we could also collect annotated information about these 559 sources (details follow), which we used as the ground truth for evaluating our methods.

5.2 Constructing input matrices

The proposed technique requires four input matrices namely

1. Social link matrix A of size $n \times n$
2. Bipartite content matrix C of size $n \times m$
3. User affinity matrix W_u and its graph Laplacian L_u

4. User affinity matrix W_v and its graph Laplacian L_v

where n is the number of users and m is the number of sources.

5.2.1 Bipartite content graph and adjacency matrix C

We use the collection of tweets and the set of users to construct a bipartite content matrix C (user $\times$ source). In order to do so, for each tweet we extract the URL (if there exists a URL) mentioned in the tweet. We then parse the URL to extract the source (news media channel) that the URL belongs to (source of the content). Next, for each *(user,source)* pair, we create an entry in the content matrix C such that the (i, j)-th element of such adjacency matrix C is equal to the number of times i-th user has tweeted/re-tweeted content from j-th source. So the size of C for our dataset is 6391×559

5.2.2 Social link graph and adjacency matrix A

For this set of 6391 users we also build their social relationship graph. For our experiments we built two variants of social graphs collected from two difference sources of information (i) *re-tweet* and (ii) *follows* on twitter.

1. directed Re-tweet graph and the corresponding non-symmetric link matrix
2. undirected follows graph and the corresponding symmetric link matrix

5.2.3 Laplacian of affinity graphs L_u and L_v

We construct user affinity and content affinity matrices W_u and W_v by computing the pairwise cosine similarity of row and column vectors of the content matrix C respectively. We then construct the graph Laplacian on the affinity graphs for users and content, $L_u = D_u - W_u$ and $L_u = D_v - W_v$ where $D_u = \sum_{j=1}^{N} (W_u)_{ij}$ and $D_v = \sum_{j=1}^{N} (W_v)_{ij}$.

In summary, at the end of data collection and processing we have four kinds of data inputs (i) symmetric/asymmetric Social Link matrices A of dimensions 6391×6391 (ii) bipartite content matrix C of dimensions 6391×559 and (iii) Laplacian of the user affinity graphs L_u of dimensions 6391×6391 and (iv) Laplacian of the source affinity graphs L_v of dimensions 559×559

5.3 Experimental setup

In this section we describe the experimental setup required for the proposed techniques described in Chapter 4 and discuss some practical issues in NMF algorithms.

5.3.1 Parameter setting

The weight parameters α and β influence the clustering. When $\alpha = 0$ only the similarity between sources has an influence in clustering. When $\beta = 0$ only the similarity between users has an influence in clustering. The scale of α, β control the strength of graph regularization in the over optimization problem. In our experiments we gradually adjusted the values of α and β. We observed that combining both link and content information and applying dual graph regularization has better performance. In our final experiments we chose the parameters by performing a grid search over a range of parameter values. We assume that we know the number of clusters *a priori*. For our experiments, since we know that US politics has two dominant ideologies "Liberal" and "Conservative", we set the number of clusters k as 2.

5.3.2 Initialization

In our experiments, we consider two kinds of initializations of the matrices U and V. In the first way, the initialization of the matrices U and V was performed randomly, according to a uniform distribution in $[0, 1]$. In the second way, we apply SVD to the input matrix C for discovering the initial structure of the matrices U and V. That is, learning the matrices U and V can be seen as a two step process, first we learn the initial structure by applying SVD and next we use the proposed method for the refinement. In our experiments we observed that SVD initialization did not significantly improve the performance.

The latent factor matrices H_1 and H_2 that capture the association between user and source clusters are initialized as identity matrices of size k. Such an initialization helped us to achieve a one-to-one block diagonal correspondence between the two types of clusterings such that $i-$th user cluster be corresponding to $i-$th source cluster.

5.3.3 Summary of the experimental steps

The matrices A, C, L_u and L_v are given as input to Equation (3). We initialize the latent factor matrices U and V using two variants of initializations (i) SVD (ii) random. The latent factors H_1 and H_2 that capture the association between user and source clusters are initialized as identity matrices of size k. We set the number of ideology clusters to be 2 manually since we have the domain knowledge for the US elections dataset. We then apply Equation (22) to update the latent variables iteratively until convergence using the multiplicative update rules in Equation (25). For each run we separate a small portion of data as validation set. During the training we apply grid search to choose the best parameter setting. We repeat the experiments five times with different starting points (initializations of latent variables). We choose the run with best performance. We normalize the latent factors U, V, H_1 and H_2 as described by Li and Ding [24]. At the end of this step we have derived user and content latent factors with a probabilistic meaning.

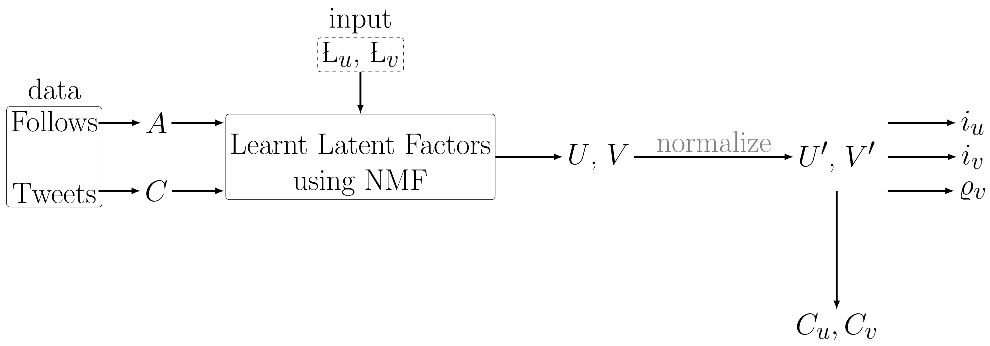


Figure 4: Experimental Steps

5.3.4 Seperating users and sources by ideology

Finally, the user and source ideological communities C_u and C_v are derived as $C_{u_i} = \arg\max U_{ij}$ and $C_{v_i} = \arg\max V_{ij}$ respectively. Table 2 illustrates some popular twitter accounts and news media channels in each ideological community.

Table 2: The results of applying the proposed technique to separate users and sources into ideology clusters (setting $k = 2$, parameters α and β chosen by Grid Search). Values in the table are representative popular accounts on twitter and popular news media outlets.

(a) User Clustering

Liberal cluster	Conservative cluster
barackobama	tedcruz
berniesanders	seanhannity
thedemocrats	gop
housedemocrats	housegop
hillaryclinton	glennbeck
senatedems	davidlimbaughs

(b) Source Clustering

Liberal cluster	Conservative cluster
washingtonpost	breitbart
nytimes	foxnews
thehill	dailycaller
huffingtonpost	dailymail
politico	washingtonexaminer
theguardian	thegatewaypundit

5.3.5 Computing ideological scores

As discussed, the main goal of this thesis is not to meagrely seperate the user and sources into ideological communities but to estimate their ideological positioning on a scale in the range $[0, 1]$, 0 being liberal extreme and 1 conservative extreme. In order to compute such a score we represent each user and source as a posterior probability vector $\langle \alpha_i, \beta_i \rangle$ and compute the user ideology score i_u, source ideology score i_v and source popularity score ϱ_v as defined in Equation (29) and Equation (30).

Refer to Figure 4 for a summary of all the experimental steps discussed in this section.

5.4 Evaluation

We evaluate the performance of the proposed method by performing an extensive comparison with well-known NMF based community detection algorithms. The focus our evaluation is twofold:

- A: evaluate the user and source ideological cluster separation
- B: evaluate the estimated ideological and popularity scores

One can argue that B is a subset of A. However, since the procedure to compute the ideology score is a new approach proposed in this thesis, we wanted to separate the two. Hence, in order to objectively validate each method in isolation we run two types of evaluation experiments.

It is noteworthy that the key goal of this thesis is to be able to compute a continuous ideology score (RQ2). While we do evaluate the ability to discover clean ideological clusters, it is a comparatively easy and well studied task. On the other hand, the process of computing a continuous ideology score is more difficult problem and not well-studied problem. As such, the comparison with other competitive clustering approaches is out of the scope of this thesis. However, for the sake of completion and easy of implementation we include a popular graph Partitioning (METIS) Karypis and Kumar [22].

5.4.1 Baseline algorithms

We compare proposed method with three types of community detection methods, i.e., relation-only, content-only, combination of relation and content. As discussed earlier our focus is on NMF based methods since these can be used to compute an ideology score. The methods are introduced as follows:

- Relation-only (R):
  - Symmetric NMF (NMFSymm): This approach is described in Ding et al. [12]. It is a 3-factor NMF on user-user and source-source similarity matrices. As it based only on symmetric relationship between rows and columns of same data type, we can only learn one clustering at a time. It is not possible to learn correspondence between the user clusters and source clusters with this method. This method is shown to be equivalent to Normalized Cut spectral clustering [11] [37].

$$\min_{U \geq 0, H \geq 0, V \geq 0} \|X - UHU^T\|_F^2, \text{ s.t. } UU^T = I \tag{31}$$

where $X = CC^T$ for user clustering and $X = C^TC$ for source clustering.

– Graph Partitioning (METIS): This is a well-known graph partitioning approach described in Karypis and Kumar [22]. As it based on two separate graph partitions on two different graphs, i.e., user-user graph (CC^T) and source-source graph (C^TC), we do not have any correspondence between the user and source clusters. Further, since it is a hard clustering algorithm it is not possible to derive continuous scores with this approach.

- Content-only (C):

  – Orthogonal NMF Tri-Factorization (ONMTF): This method is a co-clustering approach described in [12]. It is a 3-factor nonnegative matrix factorization with orthogonality constraints. It solely uses content matrix.

$$\min_{U\geq 0, H\geq 0, V\geq 0} \|C - UHV^T\|_F^2 \text{ , s.t. } UU^T = I, VV^T = I$$

  – Co-clustering with Graph Regularization (DMCC): This is a dual manifold co-clustering approach proposed in Cai et al. [9] and [20]. In this approach, in order to retain the geometric structure of graphs based on manifold assumption, we apply graph regularization constraint on both rows and columns of the content matrix. It solely uses content matrix.

$$\min_{U\geq 0, H\geq 0, V\geq 0} \|C - UHV^T\|_F^2 + \alpha \cdot \mathrm{tr}(U^T L^u U) + \beta \cdot \mathrm{tr}(V^T L^v V)$$

- Combination of relation and content (C+R):

  – Joint Link-Content Matrix Factorization (Link-Content NMF): This is a joint matrix factorization approach inspired by the formulation described in section 3 of Zhu et al. [42]. It combines both content and link matrices and performs a joint matrix factorization with shared latent variables. In this approach we only add orthogonality constraints.

$$\min_{U\geq 0, H_1\geq 0, H_2\geq 0, V\geq 0} \|A - UH_1U^T\|_F^2 + \|C - UH_2V^T\|_F^2 \text{ s.t. } UU^T = I, VV^T = I$$

  – Proposed Method: This is our approach discussed in Section 3.6. It is a joint matrix tri-factorization on both content and link matrices with orthogonality and graph regularization constraints.

$$\begin{aligned}
&\min_{U\geq 0, H1\geq 0, H3\geq 0, V\geq 0} \|A - UH_1U^T\|_F^2 + \|C - UH_3V^T\|_F^2 \\
&+ \alpha \cdot \mathrm{tr}(U^T L^u U) + \beta \cdot \mathrm{tr}(V^T L^v V) \\
&\text{s.t. } UU^T = I, VV^T = I
\end{aligned}$$

5.4.2 Ground truth construction

Ideology ground truth: We collect the ground truth for sources (News Media Channels) from multiple studies in the literature [7][14][18]. The work by Bakshy et al. [7] provided the largest set of media channel ground truth. They provide a score between -1 and 1 for 500 most shared news domains on Facebook. Flaxman et al. [14] provided the ground truth score for 100 most visited domains in Bing tool bar. We collected ground truth for 27 domains from offline survey and webpage visit data in [18]. We mapped all the scores between 0 (liberal) and 1 (conservative) which roughly measures the fraction of views/shares/clicks by a conservative user. We hand cleaned the collected list and removed obvious errors (e.g., wikipedia.org, amazon.com). For the domains which listed in multiple lists we computed the average scores. The overlapping domains from papers [14] and [18] have been shown to have a high correlation with the scores produced by Bakshy et al. [7].

In order to collect ideological ground truth for Twitter users, we use the estimated ideological score by Barberá et al. [8] using bayesian ideal point estimate as the ground truth for ideology of Twitter users. Barberá et al. [8] perform an extensive study for nearly 12 million twitter users on 12 political (e.g., 2012 presidential election, 2013 government shutdown) and nonpolitical issues (e.g., 2014 Super Bowl). The scores collected from the work by Barberá et al. [8] are mapped to the range $[0, 1]$.

In total, we collect the ground truth for 559 news media outlets and 6391 users.

Popularity ground truth: We use the aggregated number of tweets for each news media channel in the collected data set as a proxy for the popularity of the source. We normalize the number of tweets to derive a popularity score in the range $[0, 1]$

Since the collection of users is a random set of people on user, we do not have any ground truth for popularity of twitter users.

5.4.3 Evaluation measures

In order to evaluate the clustering performance we measure purity, Adjusted Rand Index (ARI), Adjusted Mutual Information (AMI) and Normalised Mutual Information (NMI). Since the ground truth as well as output of the experimental step is an ideological score in the range $[0, 1]$, we threshold the score at 0.5 to derive two ideological clusterings. Let $C = \{C_i, \ldots, C_k\}$ be the k clusters detected by the algorithm, $G = \{G_i, \ldots, G_k\}$ be the set of ground truth communities, and n be the number of data points we can define the measures as follows

- Purity is a measure of extent to which each cluster contains data points primarily from a single class [41]. In order to compute purity, for each cluster we count the number of data points from the most common class in the said cluster. Then purity is measured as the accumulated average over all the data points. Formally, *Purity* can be defined as:

$$Purity = \frac{1}{n} \sum_{i=1}^{k} \max_j |C_i \cap G_j| \tag{32}$$

  The value of purity is between $[0, 1]$.

- Rand Index is defined as the number of pairs of data points which are both located in the same predicted clusters C and the same ground truth class G, or both in different cluster and different ground truth class [32]. It penalizes both false positive and false negative decisions during clustering. Adjusted Rand Index (ARI) adjusts Rand Index to be in the range $[0, 1]$ where values close to 0 represent random cluster labelling and 1.0 represents exactly identical clusterings.

- Adjusted Mutual Information (AMI) [36] is based on the mutual information between two sets of clustering, i.e., C and G. It allows us to make the trade off between the quality of the clustering against the number of clusters by accounting for the fact that mutual information is higher for two clusterings with larger number of clusters.

In order to validate the computed ideology scores we measure correlation coefficient between the computed ideology scores, popularity score and their corresponding ground truth.

- Pearson product-moment correlation coefficients (PMCC) is measure of the linear correlation between two variables X and Y. It has a value in between $[-1, 1]$, where positive value means positive correlation, negative value means negative correlation and 0 means no correlation.

5.5 Results

5.5.1 Evaluation of clustering and ideology scores

Since the relation-only baselines, i.e, METIS and NMF Symm only perform one-side clustering or clustering one data type at a time. We compute the results by performing two separate runs for each type data type. As such, the results of relation-only methods can only be used in isolation for users and sources seperately since we do not know the correspondence between the two clusterings. The graph partioning algorthm METIS only returns the graph partiotions as output, hence it is not possible to compute ideological scores for this method.

The results of the clustering evaluation are shown as Purity, Adjusted Rand Index (ARI), Adjusted Mutual Information (AMI) and Normalized mutual information (NMI) in Table 3. In order to evaluate the ideology and popularity scores we compute Pearson correlation coefficient (R) between the computed scores and ground truth discussed in Section 5.4.2. The correlation coefficient for ideology is represented as $PMCC_i$ and correlation coefficient for popularity is represented as $PMCC_\varrho$. Since, we do not have ground truth for popularity of users we do not compute $PMCC_\varrho$ for users. Following are some noteworthy observations:

1. The proposed method has the best performance among all the methods for both user as well as source clustering. We observe that combining both link and content information gives the considerably better results than content only methods. For example, Purity of clustering for combined methods is

20% higher than content-only methods for users and 27% higher for source clustering.

2. The presence of graph regularization constraints only had a small effect on the results. For example, the results of Proposed Method are quite similar to Link-Content NMF, similarly ONMTF results are very similar to DMCC.

3. Interestingly NMF Symm, a relation only method, performs quite well for user clustering, whereas the results are very bad for source. We believe this can be explained by the input link graph to each of the algorithms. NMF Symm clustering for users takes the follows adjacency matrix as input. It is well known that the user's follows information is a good indicator of ideological stance which is reflected in the results from NMF symm. However, the input to source factorization $X = C^T \cdot C$ which captures the number of common users shared by two sources. X has noise due to various reasons (i) users might have limited number of favorite sources which they consume, as a consequence even though two sources are similar the it is not strongly reflected in the X (ii) dataset collected is random 1% sampling which might have missed many of the tweets from common sources (iii) diversity of topical interests of users affects the tweets. However, METIS, another relation only method which has the same input has performance comparable to proposed method.

In summary, the proposed method is the most suitable of all the baseline methods for the following reasons:

1. It has the best performance among all the methods in terms of all the evaluation measures. Especially, considerably higher performance than content only methods such as ONMTF and DMCC.

2. Unlike some of the baselines (e.g., NMF Symm and METIS), it is possible to compute co-clustering of both user and cluster partitions in the same formulation. As well as derive the correspondence between the two clusterings.

3. Unlike some of the baselines (e.g., METIS), it is possible to compute ideology score

5.5.2 Ideology estimates of popular news media outlets

Figure 5 visualizes popular news media outlets and their position on the computed latent ideology scale. We observe that the position of the news sources is as expected: Liberal leaning news outlets nytimes, washington post, the guardian are on the left, and conservative news outlets fox news, breitbart, rushlimbaugh are on the right. While it is easy to identify the extreme left and right sources, the neutral are more difficult, prone to errors and most interesting results to tackle the information filter bubble issue, as well as to point users to neutral news channels. Interestingly, the proposed technique could successfully estimate the ideological positioning of many

(a) Twitter Users

Method	cluster-1 size	cluster-2 size	*Purity*	*ARI*	*AMI*	*NMI*	$PMCC_i$	$PMCC_\varrho$
NMF Symm	**2438**	**2200**	**0.928**	**0.733**	**0.628**	**0.629**	**0.912**	**NA**
Proposed Method*	**2256**	**2382**	**0.925**	**0.722**	**0.620**	**0.621**	**0.904**	**NA**
Link -Content NMF	**2256**	**2382**	**0.925**	**0.722**	**0.620**	**0.621**	**0.904**	**NA**
METIS	2704	2871	0.867	0.538	0.454	0.456	NA	NA
ONMTF	1034	3604	0.743	0.234	0.233	0.266	0.756	NA
DMCC	1019	3619	0.740	0.229	0.230	0.263	0.755	NA

(b) News Media Channels

Method	cluster-1 size	cluster-2 size	*Purity*	*ARI*	*AMI*	*NMI*	$PMCC_i$	$PMCC_\varrho$
Proposed Method*	**265**	**281**	**0.826**	**0.424**	**0.346**	**0.348**	**0.827**	**0.929**
Link-Content NMF	**263**	**283**	**0.822**	**0.415**	**0.339**	**0.341**	**0.813**	**0.930**
METIS	265	281	0.819	0.405	0.318	0.320	NA	NA
ONMTF	91	455	0.606	0.039	0.145	0.181	0.746	0.593
DMCC	91	455	0.606	0.039	0.145	0.181	0.746	0.592
NMF Symm	86	460	0.597	0.031	0.135	0.171	0.752	0.597

Table 3: Evaluating the learnt ideology scores and ideological cluster seperation for Twitter-users and News Media Outlets. Best results in the table for each measure are marked in bold. Measures listed in the table are : *Purity*, Adjusted Rand Index (*ARI*), Adjusted Mutual Information (*AMI*), Normalized Mutual Information (*NMI*) and Pearson Mutual Correlation Coefficient ($PMCC_i$ for ideology and $PMCC_\varrho$ for popularity respectively)

well known neutral news channels. For instance, *yahoo.com* a web news source which has no political affiliation, *Mediaite* a news source which has been a subject of debates whether it is a "conservative" or "liberal" outlet [4]. It is interesting to see that "Whitehouse.gov" has a neutral ideological positioning.

5.5.3 Distribution of users by ideology

Figure 6 visualizes a kernel density estimate (KDE) of ideology scores of all 6391 users. The solid lines represents the ideology score computed using proposed technique. For the sake of comparison we also plot the kde of ground truth of ideology scores (dashed line). The X-axis of the plot represents the ideology score and the y-axis represents the estimated density of users at each point. Following are two noteworthy observations

- (i) We observe that there is higher density of users in the extremes than in the centre of the ideology scale. This observation coincides with the kde plot for ground truth as well as with the observations in other studies. However, it is important to note that this partly due to the data selection bias. If you recollect, there have been a filters in our dataset collection process. First, we chose controversial topics on Twitter. We then further filtered to the subset

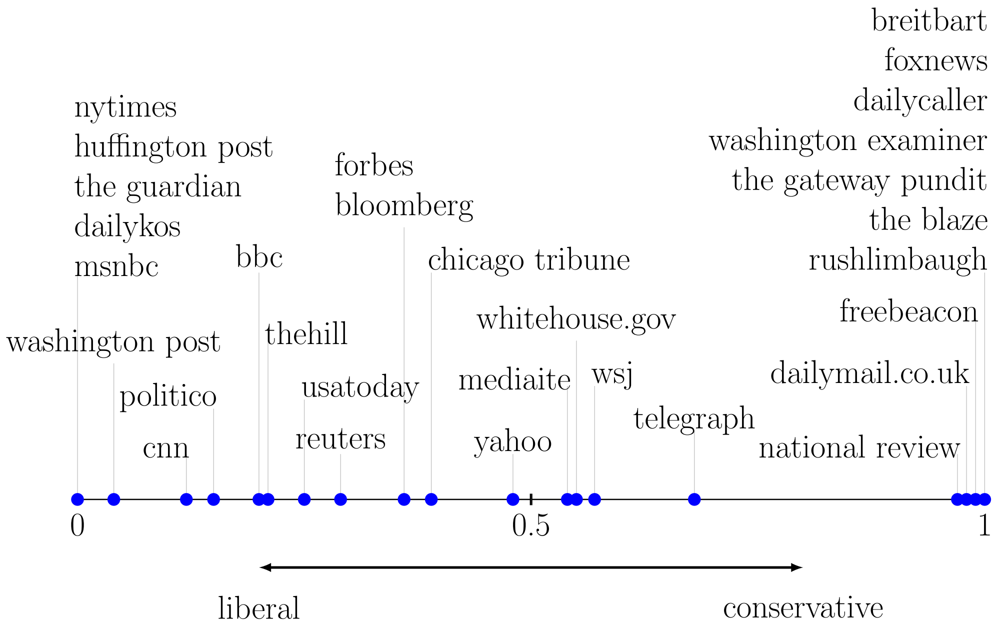


Figure 5: Popular political actors on twitter and their position on latent ideological scale

of users who tweeted on all three topics. One could assume that users who have an extreme opinion are more likely to be vocal about all the three topics. Hence, in our dataset collection we selected users who already have a strong ideology and opinion (which makes sense for our case as we want to diffuse filter bubble and this is the exact set of users who are our target) which is reflected in the plot.

- (ii) We also observe that kde of computed score has a higher skew towards the extremes when compared to the kde on ground truth, that is, the proposed technique has a tendency to exaggerate the ideology scores and shift towards the extreme. We believe this is influenced by the orthogonality constraints in the non-negative matrix factorization formulation. The orthogonality constraints in the formulation force the latent factors to be sparse. That is, if one were to visualize the two latent column vectors as axis of a latent space. We would observe the most of the latent points to be lying on the axes or close to the axes. Similarly the orthoganility constraints in the proposed formulation force the points to be close to one or the other cluster which is reflected in the plot.

5.5.4 Association between news sources and polarization of users

In order to analyze if there is any association between news sources and polarization. We bin the users according to their ideology score into 10 bins. For each bin we collect top 10 news sources. Figure 7 visualizes a binary grid with the 10 user bins on x-axis (values in the plot are average ideology score in each bin) and the top 10

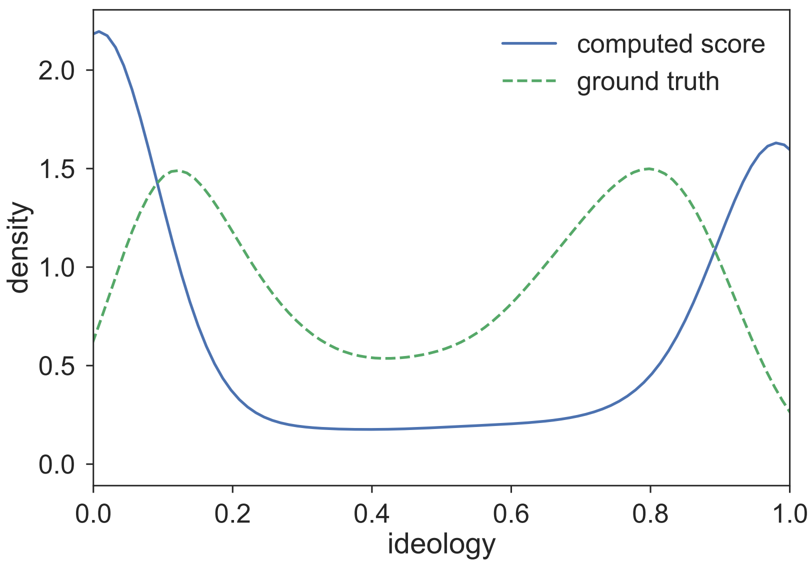


Figure 6: X axis - ideology of a user. Y axis - density of users.

popular news channels in each bin on y-axis. Both axes are sorted in the order of their ideology scores. The colored gradient in a cell in the grid can be interpreted as follows: For each dark cell in the grid the corresponding news channel on y-axis is present among the most popular 10 news sources in the corresponding user ideology bin on x-axis. For example, the cell the left top corner of the grid can be read as "*theguardian.com* news source is among the 10 most popular sources consumed by users in bin #1 with average ideology score 0.01. Following are some noteworthy observations

- Figure 7 shows that there are two clear disjoint and disconnected sets in the grid. Indicating that there is a clear, non-trivial association between preference of news sources and ideology score of users.

- *The Hill* and *Washington Post* are the two news sources which are popular across all the bins. Interestingly, *Fox News* is among the top 10 news sources for all bins except the most liberal bin (bin #1) and *nytimes* is among the top 10 news sources in all bins except the most conservative bin (bin #10). In order to simplify the plot we exclude these we exclude these four news channels from the figure.

- *The Guardian* and *Buzzfeed*, the most popular news channel in the extreme liberal bin are not popular in any other bins. Similarly, *The Blaze* and *The Gateway Pundit* are only popular in the conservative extreme.

- The users in extreme liberal bins and extreme conservative bins do not engage with sources from opposing ideology.

- As we move to the more neutral bins, ideologically extreme news sources are not popular anymore. Users in the neutral bins engage with content from neutral sources.

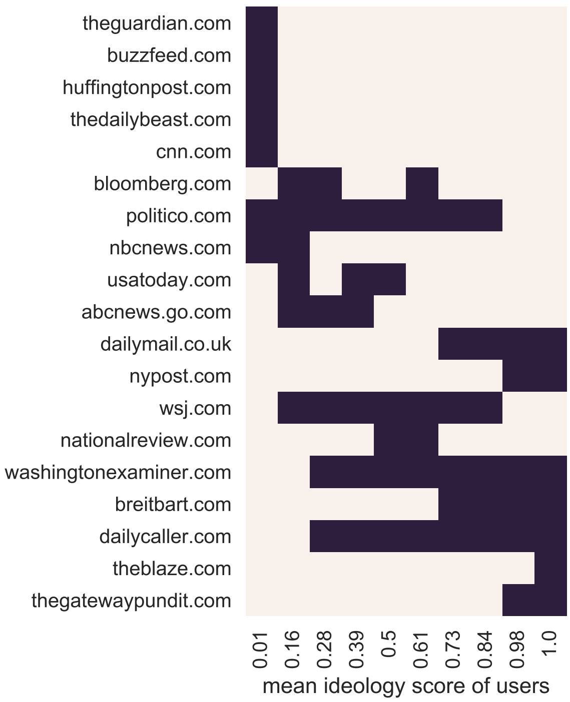


Figure 7: X axis - mean ideology of a user bin. Y axis - most popular news sources in the corresponding bin.

6 Case Study: Diffusing the Information Filter Bubble

6.1 Summary

In this section we present a case study on how the ideological scores learnt using the proposed NMF techniques can be used in two novel and interesting applications to tackle the information filter bubble. Our goal is to create solutions which can help users to explore, and hopefully reduce the information filter bubble. Unlike the previous work in this field we would like to come up with solutions that are practical, robust and fully automated. We use the algorithmic foundation built so far to develop solutions that can be easily extended and incorporated in real social network platforms and recommendation algorithms.

The contribution of this case study are two folds:

- First, we present a framework that allows users to visualize their own information filter bubble. We present a system where users can visually explore their own ideological positioning and the positioning of the content they consume.
- Second, we present preliminary work on an algorithmic fully automated approach to increase diversity of viewpoints in content that a user consumes via i) self-exploration of the content space ii)user-guided transparent recommendations.

6.2 Making users aware of their own information filter bubble

6.2.1 Motivation and Scope

The first step in diffusing a filter bubble is making the user aware of his/her information filter bubble. Recent studies have shown that making a user aware of their imbalance in reading history can encourage the users to make small but proven improvements in increasing the diversity of viewpoints in their reading [28] [19]. Inspired by the results of these studies we call for raising the awareness of social network users by providing visual, explanatory and exploratory evidence of their own information filter bubble in the form of an interactive platform.

In order to create frameworks which promote human interpretable explanations of these information filter bubbles, one approach is to visualize the user and content in the same space such that we allow users to understand their ideological leaning as well as their content bias by visual exploration of their content consumption. A simple way to do this is to define a distance measure (in some latent embedding) between user-content and content-content. We can then use these distances to map users and content to a Euclidean space.

6.2.2 Proposed Method

In this section we present a unified platform where users can explore their own ideological positioning, content ideological positioning, and their own content consumption. In order to do so, we use the shared latent space learnt using the proposed technique in chapter 3. Since the ideology scores computed using the proposed technique are in the same range $[0, 1]$ as well as relative to each other, we can use these the score to map users and sources to a euclidean space of ideology score and popularity scores. We then use principles of human perception and design [38] to enhance visual explanation of the filter bubble by including visually interpretable elements such as color, connections, distance and size to explain the information filter bubble.

Steps:

1. First, we combine user's social link information and their content consumption information using the proposed NMF based latent space learning described in chapter 3).

2. Next, we estimate the ideological positioning of users and content using the learnt latent space as described in section 4.3.

3. Using their estimated ideological positioning in step 2 we project the user and all the content related to the topic in a two dimension space. Since all user scores are on the same scale and relative to each other we would observe that user's with similar ideology are close to each other. Further, since the user and source ideology are on the same range $[0, 1]$, ideologically similar content and source will be close to each other.

4. Finally, we connect the user to the content that they have consumed. In order to increase the ease of visual interpretation, we color the content according to the ideological learning (blue: liberal, green: neutral and red: conservative). Content not consumed by the user is grey in color. The size of a the circle for a source is proportionate to the number of times a user has consumed content from the said source.

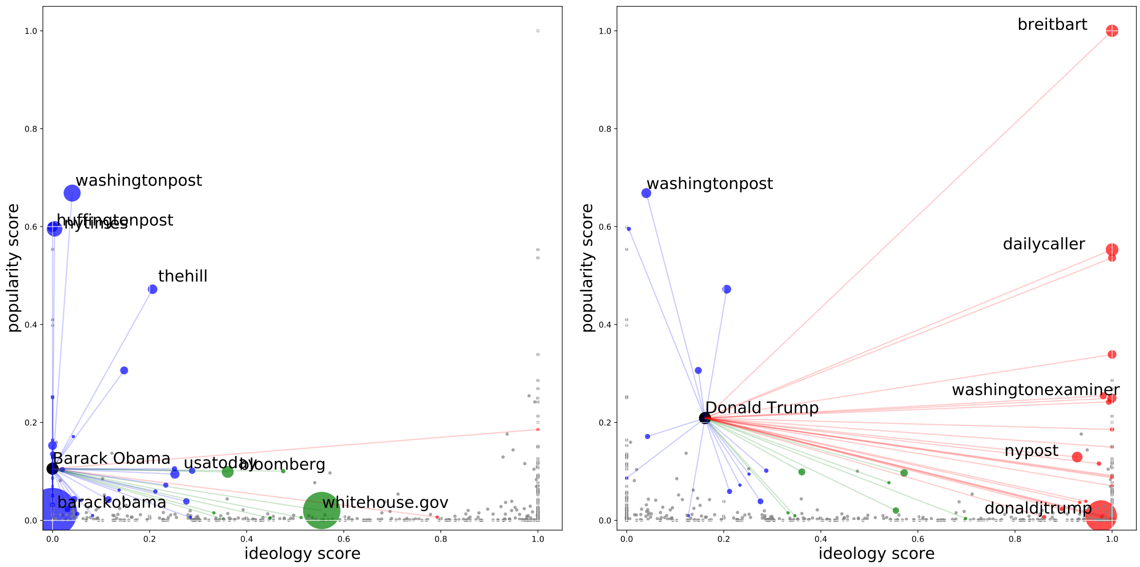
6.2.3 Prototype

Figure 8 presents a prototype for two popular twitter users - Barack Obama (@barack-obama) and Donald Trump(@realdonaldtrump). We have ideology score on x-axis (0 being liberal and 1 conservative extreme) and popularity score of sources on Y-axis. The position of a user on the space is determined by his own ideology score and the average popularity score of the content that the user engages with. The position of a source on the space is determined by ideology score and popularity score of the content respectively.

From this figure one can visually observe their own ideological positioning as well the ideology of the content that they engage with. For instance, Barack Obama is heavily liberal in his ideology (ideology score 0). The content consumed by Barack

Obama is also heavily biased on the liberal side. As expected the highest content he engages with is from *barackobama.com*. He consumes negligible amount of content from the opposite point of view. It is interesting to see that *whitehouse.gov* is a neutral source (ideology score 0.5).

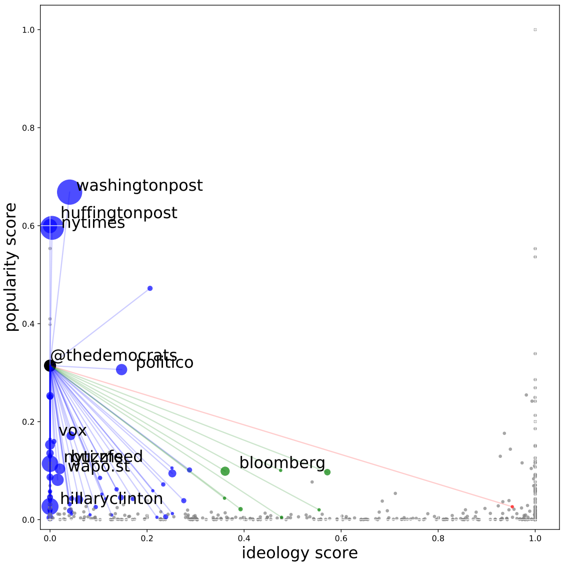
On the other hand, Donald Trump heavily engages with content on the conservative side, highest being *donaldtrump.com*. As opposed to Barack Obama, Donald Trump has a higher engagement with content from diverse view points. Interestingly, we observe that, in spite of the highly conservative content consumption the ideological positioning estimated for Donald Trump is leaning towards the liberal side. We believe this can be attributed to the mixing with the user-user matrix. Perhaps, due to his popularity users from the opposing viewpoint heavily engage with Donald Trump in spite of difference in ideology. However, this engagement need to mean endorsement. It could be attributed to, for example, parody tweets.



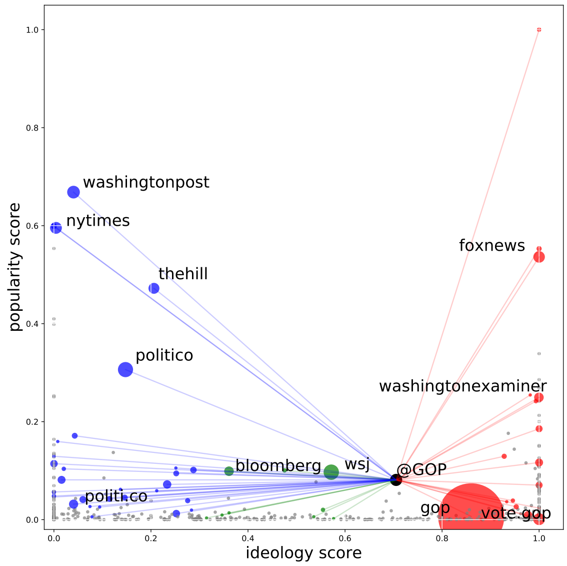
(a) official Twitter account of Barack Obama **(b)** official Twitter account of Donald Trump

Figure 8: ideological position of a twitter user and his content engagement. Points in the grey are the sources that the user never interacted with. Note: only the sources are tweeted more than 10 times are labelled.

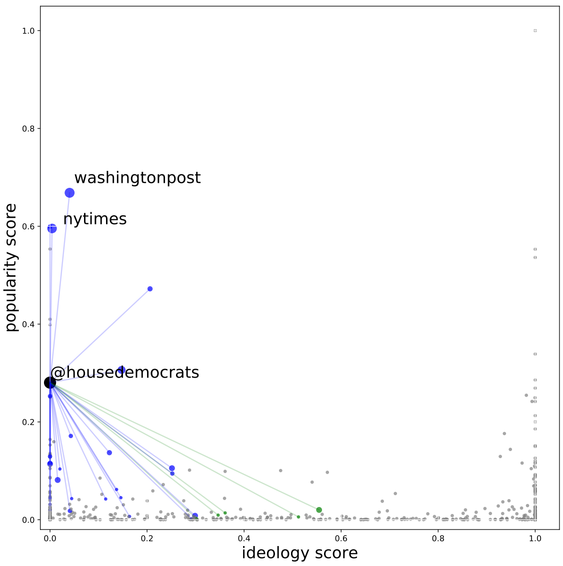
Figure 9 presents more examples of some popular twitter accounts. For the purpose of illustration we have selected accounts (official Twitter accounts of republican party, democratic party, and the republican and democratic house of representatives) for which ideological leaning is obvious due to their political association. We can observe that once again the users have their own information filter bubbles. It is interesting to observe that republican accounts have higher engagement with diverse view points than liberals. This could also be attributed to the time period in which the tweets were collected (in the period 2011 to 2016 when the tweets were collected United States had a liberal President).



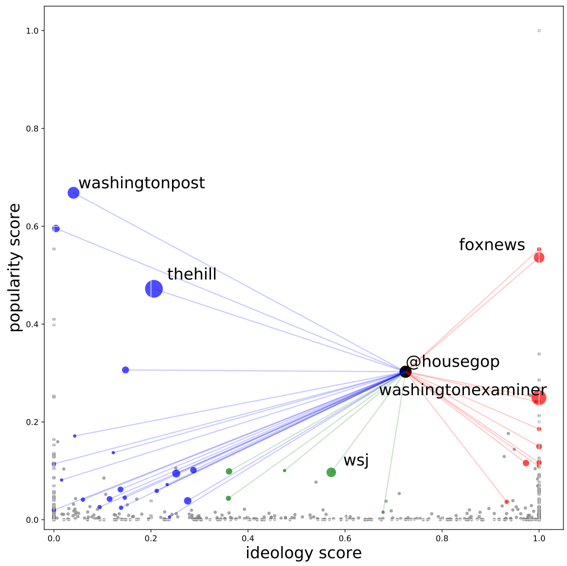
(a) official Twitter account of the Democratic Party



(b) official Twitter account of the Republican Party



(c) official account of House Democrats



(d) official account of House Republicans

Figure 9: ideological position of a twitter user and his content engagement. Points in the grey are the sources that the user never interacted with. Note: only the sources are tweeted more than 10 times are labelled.

6.3 Diversify user's content-consumption using recommendation

6.3.1 Motivation and Scope

In chapter 1 we discussed the merits of connecting a user to content from opposing view point, and the role that such an exposure to diverse view points would play in reducing the filter bubble. In this section we discuss how content recommendation can be used to diversify user's content consumption, consequently reduce the information filter bubble.

Our main motivation comes from the recent work by Garimella et al. [17] on connecting a user with opposing views by recommending content from opposite side. They suggest that a "successful" recommendation from opposing view point can have the highest impact on reducing polarization of users. In this context *successful* would refer to a recommendation which is likely to be accepted by user. Clearly, users have a tolerance level and are more likely to accept content from the opposing point of view until a certain degree. For instance, it is unlikely that an extremely conservative view point would be successfully accepted (viewed) and endorsed(tweeted/re-tweeted) by a liberal user. Whereas, content which is ideologically opposing but with a certain tolerance distance is much more likely to be accepted. Further, Garimella et al. [17] suggest that highest reduction in user-polarization score is achieved by connecting a user with authoritative source with the opposing view. That is, maximum reduction in user-polarization happens when a user is recommended "popular" content from the opposing viewpoint. The intuition behind the idea is that in this way the user can see the "popular" notion on the other side, and presumably the "popular" content is usually of good quality.

We use this as the basis for our research in building prototypes that encourage engagement with content from diverse viewpoints. We present preliminary work on how the latent factors learnt in the proposed technique in Section 4.3 can be used to recommend content to a user. However, a good recommendation usually aims to suggest content that is likely to be consumed by user, i.e., recommend content which is similar to topical interests of user as well as adheres to ideological leaning of a user. This recommendation, by definition reinforces the information filter bubble. Hence, for the purpose of diffusing filter bubble we restate the recommendation problem as the problem of - "recommending content that is similar to user's topical interest but diverse from a user's ideological viewpoint". Thus, the proposed recommendation takes into account three factors - "user's topical interest", "ideology of a source" and "popularity of a source".

6.3.2 User and Content Topic Modelling

In order to make content recommendations by topic, we first need to model user as well as content in a *topic latent space*. Given N documents in a collection, we can construct a document-term matrix $D \in \mathbb{R}^{N \times T}$ where T is the set of term features extracted from the documents. D_{ij} corresponds to the significance of term t_j in document d_i which can be calculated as

$$D_{ij} = TF_{ij} \log(\frac{N}{DF_j}) \tag{33}$$

where TF_{ij} denotes the frequency of the term t_j in document d_i and DF_j is the number of documents in the collection in which the term t_j appears. D_{ij} is a non-negative matrix.

In our Twitter dataset we have 33600 documents which are divided into 4 categories: obamacare, abortion, gun control, and others (noise during dataset collection). For each URL in our twitter dataset collection, we parse the URL and extract the content from the web page. We build a tf-idf model on the 33600 parsed documents. We select the top 100 term features and build a *document-term* matrix D of dimension 33600×100 as described in Equation 33.

Next, in order to learn co-clustering of documents and terms by topic. We run the orthogonal non-negative matrix tri-factorization (ONMTF) algorithm for co-clustering described in [12] on input matrix D. We run the algorithm under different parameter settings using grid search and choose the best result. The ONMTF algorithm factorizes the document-term matrix D as follows:

$$\min_{U \geq 0, H \geq 0, V \geq 0} \|D - VHW^T\|_F^2 \tag{34}$$

$$\text{s.t. } VV^T = I, WW^T = I \tag{35}$$

where V captures the document-topic clustering and W captures the term-topic clustering. We set the number of clusters to 4 as we know the true number of topical categories in our dataset. Each document v_{ij} in V can be seen as the probability of $i-$th document belonging to the $j-$th topic cluster. Similarly, term w_{ij} in W can be seen as the probability of $i-$th term belonging to the $j-$th topic cluster. That is we have soft clustering for both documents and terms. Figure 10 visualizes the top ten features in each column vector of the term factorization component W. The extent of color fill in each cell represents the degree to which the said term belongs to the cluster. We validate the term clustering against manually labeled ground truth for the most common term features. The algorithm separates the terms in the three topics - obamacare, gun control and abortion with 100% accuracy. In order to derive topic interests of users we refer to document topic clustering and assign topical interests to users as the fraction of documents consumed in each topic. We evaluate URL-topic clusterings by randomly selecting a sample of documents and manually validating. We do not perform any quantitative analysis since this is not the main focus of the thesis.

6.3.3 Examples: ideological position of content for each topic

For each topic we randomly select a few URLs positioned at left, right and centre of the ideology scale and visualize them. Figure 11, 12, and 13 visualize some examples of content for each topic and their ideological position.

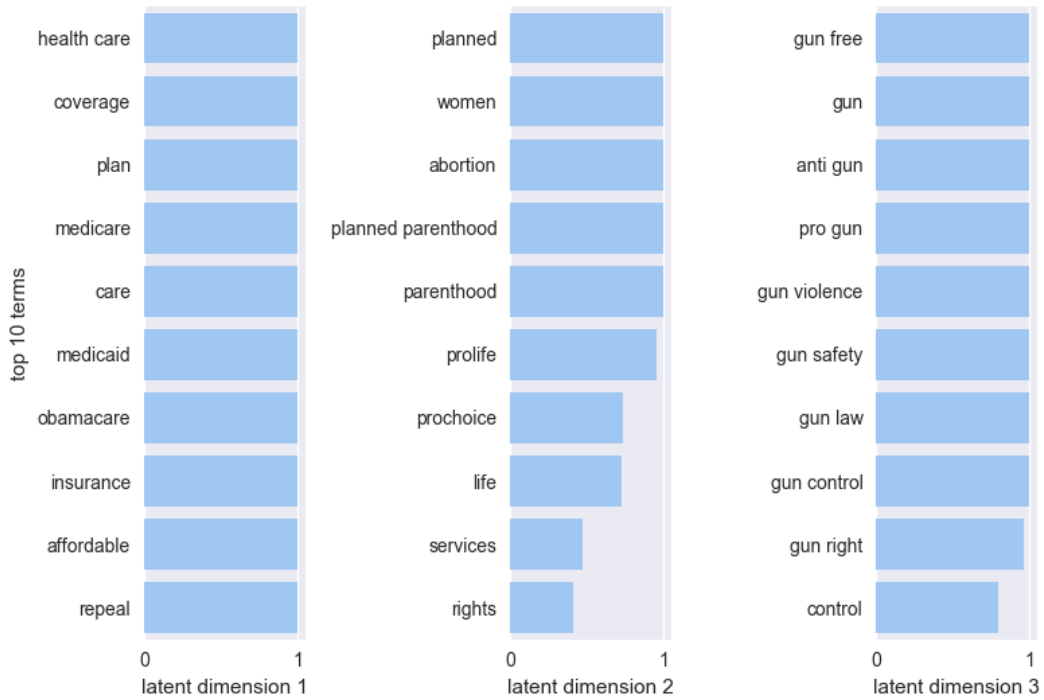


Figure 10: top 10 term features for each topical latent vector

Obamacare: Figure 11 displays few randomly selected representative sample of URLs related to the topic "Obamacare" and their position on the ideological latent space. As expected we observe that the content on the centre of the ideological scale is close to being neutral in the view point. The language used in the URL text is straightforward and objective. As we move towards either extreme ends we can observe that the content becomes more and more extreme in viewpoint, which to some extent is also reflected in the language and words used in the URL text. We can see that the as we move towards extremes even the language in the URLs (e.g., "obamacare is a failed product") becomes similar to the language used by click baits.

Gun control: Figure 12 displays few randomly selected representative sample of URLs related to the topic "Gun control" and their position on the ideological latent space. It is interesting to see that once again there is a clear difference in viewpoints on either side of the scale. The notion of "gun violence" is more prevalent in the liberal leaning content whereas the terms "gun rights" "gun freedom" are more popular in the conservative leaning content. As expected we observe that the content on the centre of the ideological scale is close to being neutral in the view point meagerly stating facts.

Abortion: Figure 13 displays few randomly selected representative sample of URLs related to the topic "abortion" and their position on the ideological latent space. As expected, again there is a clear difference in viewpoints on either side of the scale. The views "women rights", "choice", 'planned parenthood" and "legalizing abortion" is more prevalent in the liberal leaning content. Whereas the conservative leaning

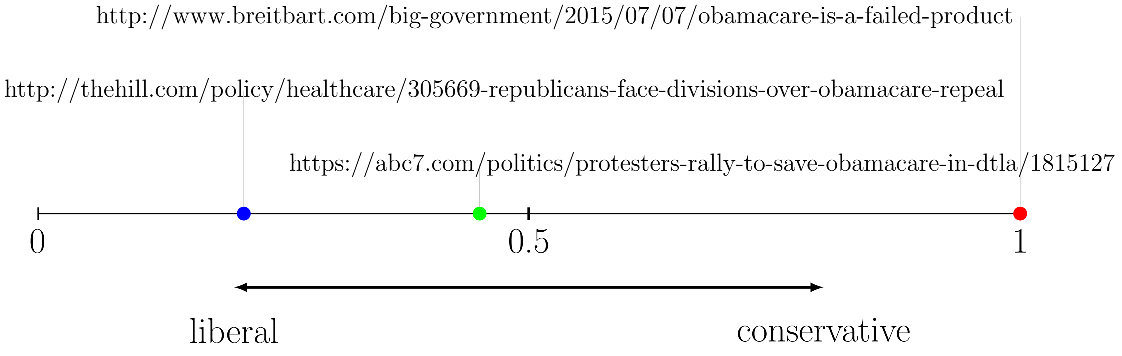


Figure 11: Randomly selected urls for the topic "obamacare" and their position on latent ideological scale

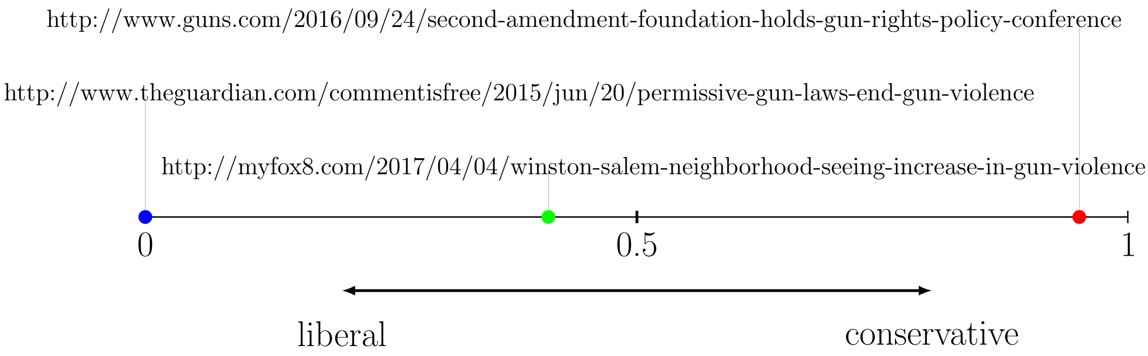


Figure 12: Randomly selected urls for the topic "guncontrol" and their position on latent ideological scale

content strongly apposes these views. Once again, the content close to the centre of the scale is neutral in perspective. For example, the article posted by a neutral source *yahoo.com* is titled "Activists On Both Sides Of Abortion Issue To Protest Across U.S.".

6.3.4 Proposed Method and Prototype

Steps:

1. First, we compute topical clustering on content and estimate the user's topical interest as described in sec 6.3.2.

2. We compute the ideology scores and popularity scores as described in Chapter 4. We then project the user as well as the content on the same latent space. Figure 14a visualizes a hypothetical user in the original ideology latent space and the transformed space in ideology-popularity dimensions. As discussed earlier, each user has an ideological tolerance θ. A user is more likely to accept

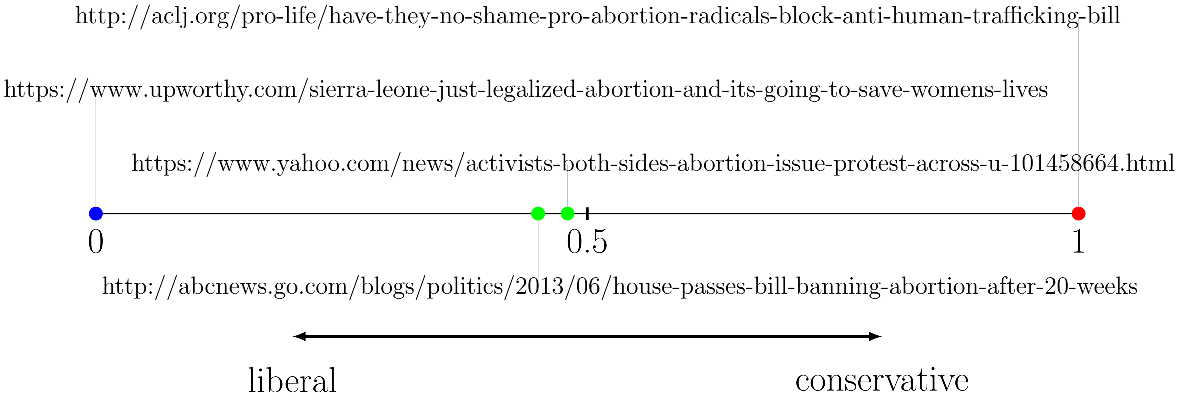


Figure 13: Randomly selected urls for the topic "abortion" and their position on latent ideological scale

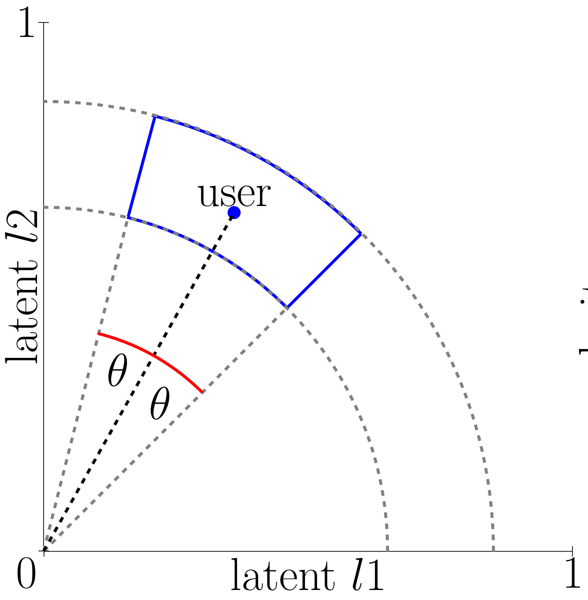
content within the region $+\theta$ and $-\theta$ on either side of the user's ideological positioning.

3. For a better understanding of the user and content position, we transform the user and all the content to the corresponding ideology-popularity coordinate space in Figure 14b.

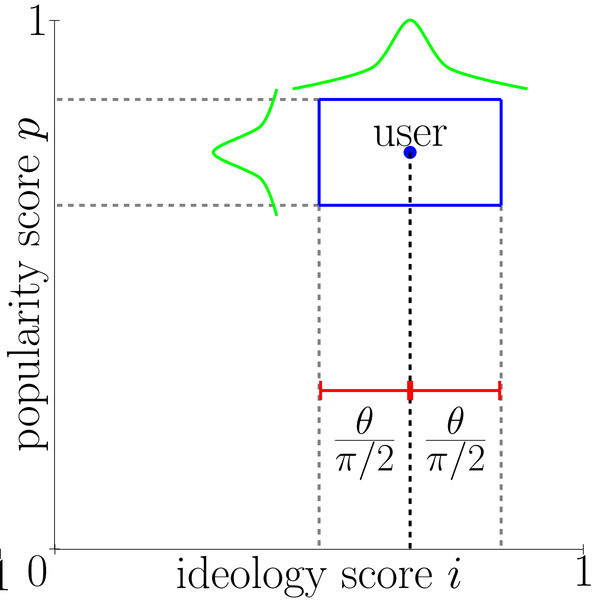
4. Consider that we build two Gaussians over around the user box (parallel to x and y axes) with their means centered at user's ideology and popularity score respectively, and variance as a function of the tolerance threshold given as input by the user. We can now sample content from these gaussians distributions and use it for recommending content to the user. As desired, in such a sampling, the content close to the user's own ideology and popularity score has a higher probability of being selected, and as we move towards the thresholds the probability of an article being selected gradually decreases (and eventually becomes practically zero).

6.3.5 Future extension - Next Generation Social Networks

The system discussed in this section can be easily extended to build an interactive and exploratory interfaces that allow the users to visually explore the content related to a topic and self-adjust diversity of their content consumption. For instance, given a topic (e.g., Gun control), we can visualize the user (his viewpoint), content that he creates/consumes (content filtered in), and representative examples of other viewpoints (content filtered out) in the same latent space. Visualizing all the diverse viewpoints (e.g., representative content) relevant to the topic and their distance from the user, allows user to visually understand the self-user bias as well as the bias imposed by the algorithmic filters. We believe that such systems will be of high value in designing, what we call, *Next Generation Social Networks* - Futuristic social media platforms that encourage discussion and debates between users of diverse view



(a) user in original latent space



(b) user in transformed latent space

Figure 14: logical diagram of user content recommendation by sampling from the gaussian over "ideology" and "popularity" positioning

points; Social networks that help reducing the ideological segregation of users instead of reinforcing it.

7 Discussion

7.1 Conclusions

In this thesis we presented a framework to uncover ideological latent space in Twitter, and show that highly accurate clustering of users and sources (according to their ideological stance), as well as an ideology score can be recovered from Twitter data. We combine information from user's social link data and content consumption data. Using matrix factorization on the joint formulation with shared latent variables we uncover a common ideological latent space of users and sources on Twitter.

Our framework distinguishes itself from most existing work in filter bubble in two major ways. First, our model aims to learn ideology on a continuous scale rather than a binary *liberal-conservative* opinion which is much simpler and well studied problem. Second, our model defines filter bubble as a multidimensional problem, and allows for learning any number of dimensions rather than just one (ideology). These elements give our model explanatory power that can be used to not only diffuse the information filter bubble but also to use it as a toolkit to address a variety of social phenomenon namely, intolerance to opposing viewpoints, conformation bias, polarization of views, and ideological segregation.

The collective and simultaneous matrix factorization on both link and content information to learn a shared ideological model for both user and source gave our proposed technique valuable explanatory observation powers. For instance, we could identify a user's ideological stance on a topic and their association with news sources in a unified unsupervised approach automatically (with no content/sentiment analysis and natural language processing).

Looking at the information filter bubble, as a problem of learning multidimensional latent space proved to be fruitful as well. For example, while we embarked on the quest to learn ideological latent space, we serendipitously discovered the dimension "popularity" of the source, which is an interesting dimension to address the filter bubble effect. In our experiments, we show that indeed the discovered has a high correlation with the ground truth for popularity.

Finally, we present a prototype to demonstrate how we can create interactive interfaces using the user's and content's ideology score which visualize the filter bubble. We also explain how we can use the score to make recommendations that help users diversify their news feed, consequently burst this filter bubble.

7.2 Implications to future research

Lessons from this framework could be applied to another parallel line of futuristic work related to "interpretable and explanatory machine learning and data mining". In the recent times there has been a growing concern over algorithmic decision making. Machine learning based systems, for instance, online recommendation systems act like black boxes, often making decisions on behalf of users, without informing them, thus leaving them unaware of personalization employed as well as various choices available. In recent times, many online platforms have attempted to explain the decision making humanly. For example, to explain their content based recommendations, Netflix offers - "because you watched". In order to explain their collaborative filtering, Amazon offers - "customers who bought xxx, also bought yyy". While these are steps taken in the right direction, there are still unaddressed gaps. For instance, both of these examples focus on explaining what was `filtered IN` by the algorithm. However, it is equally or in fact more important that the user knows what was `filtered OUT`. We believe the case studies presented in chapter 6 will motivate the reader to extend the model in this direction.

Further, the system discussed in chapter 6 can be easily extended to build futuristic social network that allow the users to visually explore the content related to a topic and self-adjust diversity of their content consumption. We believe such systems will be of high value in designing social media platforms which encourage discussion and debates between users of diverse view points. We call such social network as *Next Generation Social Networks* - "social networks which help reduce the ideological segregation of users instead of reinforcing them".

Lastly, I believe, this research is a small step towards broader issues in ethics of big data that are in rise, and would continue to rise in the next few years. To give an example, European Union's General Data Protection Regulation (GDPR 2018) [1] covers two key aspects of data and algorithms - "algorithmic decision-making" and "right to explanation". The GDPR policy on "Right to explanation" would require algorithms to provide user an explanation for algorithmic decisions made for them. While this policy poses large challenges for the industry, it highlights the gap between the legal aspirations and technical realities. In its current form, there is a large disparity between the complexity of algorithms and desired legal frameworks. Perhaps some of the results of this research could be used to address reduce this gaps, as well as guide technology and industry, for example, in adapting the GDPR policy if and when it becomes a reality in 2018.

8 REFERENCES

[1] Parliament and council of the european union (2016). general data protection regulation.

[2] Blue Feed, Red Feed. accessed 25-June-2017, http://graphics.wsj.com/blue-feed-red-feed/, 2012.

[3] Escape Your Bubble. accessed 25-June-2017, https://www.escapeyourbubble.com/, 2012.

[4] so-is-mediaite-conservative-or-libera. accessed 25-June-2017, http://www.dan-abrams.com/so-is-mediaite-conservative-or-liberal/, 2012.

[5] L. Akoglu. Quantifying political polarity based on bipartite opinion networks. In *ICWSM*, 2014.

[6] J. An, D. Quercia, and J. Crowcroft. Partisan sharing: facebook evidence and societal consequences. In *Proceedings of the second ACM conference on Online social networks*, pages 13–24. ACM, 2014.

[7] E. Bakshy, S. Messing, and L. Adamic. Exposure to diverse information on facebook. *Facebook Research Blog*, 2015.

[8] P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science*, 26(10):1531–1542, 2015.

[9] D. Cai, X. He, X. Wu, and J. Han. Non-negative matrix factorization on manifold. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, pages 63–72. IEEE, 2008.

[10] M. Conover, J. Ratkiewicz, M. R. Francisco, B. Gonçalves, F. Menczer, and A. Flammini. Political polarization on twitter. *ICWSM*, 133:89–96, 2011.

[11] C. Ding, X. He, and H. D. Simon. On the equivalence of nonnegative matrix factorization and spectral clustering. In *Proceedings of the 2005 SIAM International Conference on Data Mining*, pages 606–610. SIAM, 2005.

[12] C. Ding, T. Li, W. Peng, and H. Park. Orthogonal nonnegative matrix t-factorizations for clustering. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 126–135. ACM, 2006.

[13] S. Dori-Hacohen, M. Jang, and J. Allan. Is climate change controversial? modeling controversy as contention within populations. *arXiv preprint arXiv:1703.10111*, 2017.

[14] S. Flaxman, S. Goel, and J. M. Rao. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1):298–320, 2016.

[15] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis. Quantifying controversy in social media. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 33–42. ACM, 2016.

[16] K. Garimella, G. De Francisc iMorales, A. Gionis, and M. Mathioudakis. Mary, mary, quite contrary: Exposing twitter users to contrarian news. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 201–205. International World Wide Web Conferences Steering Committee, 2017.

[17] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis. Reducing controversy by connecting opposing views. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 81–90. ACM, 2017.

[18] M. Gentzkow and J. M. Shapiro. Ideological segregation online and offline. *The*

Quarterly Journal of Economics, 126(4):1799–1839, 2011.

[19] E. Graells-Garrido, M. Lalmas, and D. Quercia. Data portraits: connecting people of opposing views. *arXiv preprint arXiv:1311.4658*, 2013.

[20] Q. Gu and J. Zhou. Co-clustering on manifolds. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 359–368. ACM, 2009.

[21] K. Hosanagar, D. Fleder, D. Lee, and A. Buja. Will the global village fracture into tribes? recommender systems and their effects on consumer fragmentation. *Management Science*, 60(4):805–823, 2013.

[22] G. Karypis and V. Kumar. Unstructured graph partitioning and sparse matrix ordering system, version 2.0, 1995.

[23] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[24] T. Li and C. Ding. The relationships among various nonnegative matrix factorization methods for clustering. In *Data Mining, 2006. ICDM'06. Sixth International Conference on*, pages 362–371. IEEE, 2006.

[25] Q. V. Liao and W.-T. Fu. Can you hear me now?: mitigating the echo chamber effect by source position indicators. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 184–196. ACM, 2014.

[26] H. Lu, J. Caverlee, and W. Niu. Biaswatch: A lightweight system for discovering and tracking topic-sensitive opinion bias in social media. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 213–222. ACM, 2015.

[27] A. Morales, J. Borondo, J. C. Losada, and R. M. Benito. Measuring political polarization: Twitter shows the two sides of venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(3):033114, 2015.

[28] S. A. Munson, S. Y. Lee, and P. Resnick. Encouraging reading of diverse political viewpoints with a browser widget. In *ICWSM*, 2013.

[29] E. Pariser. *The filter bubble: What the Internet is hiding from you.* Penguin UK, 2011.

[30] S. Park, S. Lee, and J. Song. Aspect-level news browsing: understanding news events from multiple viewpoints. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 41–50. ACM, 2010.

[31] M. Prior. *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections.* Cambridge University Press, 2007.

[32] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.

[33] P. Resnick, J. Konstan, and A. Jameson. Panel on the filter bubble. In *The 5th ACM conference on Recommender systems*, 2011.

[34] P. Resnick, R. K. Garrett, T. Kriplean, S. A. Munson, and N. J. Stroud. Bursting your (filter) bubble: strategies for promoting diverse exposure. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*, pages 95–100. ACM, 2013.

[35] C. R. Sunstein. *Republic. com 2.0.* Princeton University Press, 2009.

[36] N. X. Vinh, J. Epps, and J. Bailey. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of*

Machine Learning Research, 11(Oct):2837–2854, 2010.

[37] D. Wang, T. Li, S. Zhu, and C. Ding. Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 307–314. ACM, 2008.

[38] C. Ware. *Information visualization: perception for design.* Elsevier, 2012.

[39] F. M. F. Wong, C. W. Tan, S. Sen, and M. Chiang. Quantifying political leaning from tweets and retweets. *ICWSM*, 13:640–649, 2013.

[40] J. Yoo and S. Choi. Orthogonal nonnegative matrix tri-factorization for co-clustering: Multiplicative updates on stiefel manifolds. *Information processing & management*, 46(5):559–570, 2010.

[41] Y. Zhao and G. Karypis. Criterion functions for document clustering: Experiments and analysis. Technical report, Technical report, 2001.

[42] S. Zhu, K. Yu, Y. Chi, and Y. Gong. Combining content and link for classification using matrix factorization. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 487–494. ACM, 2007.